Local Limit Theorem in Random Graphs and Graphs on Non-Constant Surfaces

Sophia Saller
St Catherine's College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy
Trinity Term 2019

To the memory of my dad, who never stopped believing in me.

Acknowledgements

I would like to express my sincere gratitude to my supervisors Prof. Oliver Riordan and Prof. Colin McDiarmid for their continuous support, for their patience, motivation, and immense knowledge. Their guidance helped me in all the time of research and writing of this thesis. I was very lucky to have such knowledgeable, supportive and helpful supervisors.

I also want to thank my colleagues for providing a friendly and productive environment in which to do research. Further, I would like to thank all the students I had the privilege to teach for making teaching a fun and sometimes mathematically challenging experience.

Lastly, I would like to thank my family and friends for all their support, belief and patience during the writing of this thesis.

Statment of Originality

The work contained in this thesis has not been previously submitted for a degree or diploma at any other higher education institution. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due references are made.

The first Chapter is joint work with my supervisor Prof. Oliver Riordan, the second chapter is joint work with my supervisor Prof. Colin McDiarmid.

Abstract

The thesis is split into two parts. In the first part we prove a local limit theorem for the number of appearances of the complete graph on four vertices, K_4, in the Erdős–Rényi Random graph $G(n,p)$ for $p \in (0,1)$ a fixed constant. The proof of this is based on bounding the characteristic function $\psi_n(t)$ of the number of K_4 in $G(n,p)$ by using different conditioning arguments for different ranges of t. The second part treats classes of graphs embeddable in either an orientable surface of Euler genus at most $g(n)$, denoted by $\mathscr{G}_n^{g(n)}$, or non-orientable surface of Euler genus at most $g(n)$, denoted by $\mathscr{H}_n^{g(n)}$, for different non-constant functions $g(n)$. We first find bounds on the sizes of these classes and prove that as long as $g(n) = o\left(\frac{n}{\log^3(n)}\right)$, the classes $\mathscr{G}_n^{g(n)}$ and $\mathscr{H}_n^{g(n)}$ both have a growth constant equal to the planar graph growth constant. As long as $g(n) = O\left(\frac{n}{\log(n)}\right)$ it is further shown that the classes $\mathscr{G}_n^{g(n)}$ and $\mathscr{H}_n^{g(n)}$ are both small. We then go on to show which properties graphs in these graph classes commonly display, such as giving bounds on the number of edges, faces, pendant vertices, the maximum degree and the probability of connectedness.

Contents

1 Introduction **1**

1.1 Background 1

1.2 Main Ideas in Thesis 3

1.2.1 Local Limit Theorem for Number of K_4 Subgraphs in $G(n,p)$ . . 3

1.2.2 Graphs on Non-constant Surfaces 4

2 Local Limit Theorem for Number of K_4 Subgraphs in $G(n,p)$ **7**

2.1 Introduction 7

2.1.1 Graph Containment Thresholds 8

2.1.2 Central Limit Theorems 9

2.1.3 Local Limit Theorem for Triangle Count 11

2.2 Statement of Main Result 11

2.3 Overview of Proof Method 12

2.4 Fundamental Proof Strategy 12

2.5 Notation and Preliminaries 15

2.5.1 Central Limit Theorem for Number of K_4 18

2.5.2 Chernoff Bounds 19

2.6 Outline of Main Conditioning Argument 19

2.7 Proof of Main Theorem 21

2.7.1 Region $R_1 = (0, A/\sigma_n)$ 21

2.7.2 Region $R_3 = (n^{-2.45}, n^{-2.1})$ 22

2.7.2.1 Bounding $\mathbb{E}_F\left[\left|\mathbb{E}_E\left[e^{i\theta\,\mathbf{Y}(E,F)}\right]\right|\right]$ 23

2.7.2.2 Bounding $\mathbb{E}\left[\left|e^{i\theta\cdot(\mathbf{Z}(E,F)-\alpha)}-1\right|\right]$ 25

2.7.2.3 Combining Bounds 26

2.7.3 Region $R_2 = (0, n^{-2.45})$ 27

2.7.3.1 Bounding $\left|\mathbb{E}\left[e^{i\theta(\mathbf{C}(F)+\mathbf{W}(E,F))}\right]\right|$ 28

2.7.3.2 Bounding $\mathbb{E}\left[\left|e^{i\theta(\mathbf{Y}(E,F)-\alpha)}-1\right|\right]$ 29

2.7.3.3 Bounding $\mathbb{E}\left[\left|e^{i\theta(\mathbf{Z}(E,F)-\beta)}-1\right|\right]$ 31

2.7.3.4 Combining Bounds 32

2.7.4 Region $R_5 = (n^{-1.96}, n^{-1.4})$ . . . 33
2.7.5 Region $R_4 = (n^{-2.1}, n^{-1.96})$ . . . 45
2.7.6 Region $R_7 = (n^{-0.6}, \pi)$ . . . 55
2.7.7 Region $R_6 = (n^{-1.4}, n^{-0.6})$ . . . 61
2.8 Conclusion and Further Open Problems . . . 66

3 Graphs on Non-constant Surfaces 67
3.1 Introduction . . . 67
3.1.1 Properties of Planar Graphs . . . 79
3.1.2 Properties of Graphs on Surfaces . . . 82
3.2 Basic Results on Graphs on Surfaces . . . 84
3.3 The Size of Graph Classes . . . 96
3.3.1 Lower Bounds . . . 98
3.3.2 Upper Bounds . . . 102
3.3.2.1 Orientable Case . . . 103
3.3.2.2 Non-orientable Case . . . 106
3.3.2.3 Growth Constant up to Genus $g(n) = o\left(\frac{n}{\log^3(n)}\right)$ . . . 113
3.3.2.4 Alternative Proof of Growth Constant up to Genus $g(n) = o\left(\frac{n}{\log^{3+\epsilon}(n)}\right)$ . . . 119
3.4 Properties of Graph Classes . . . 127
3.4.1 Number of Edges and Number of Faces . . . 127
3.4.2 Number of Pendant Vertices . . . 131
3.4.2.1 Class of Graphs of Euler Genus $g(n) = O\left(\frac{n}{\log(n)}\right)$ . . . 132
3.4.2.2 Class of Graphs of Euler Genus $o\left(\frac{n}{\log^3(n)}\right)$ . . . 133
3.4.2.3 Class of Graphs of Euler Genus $g(n) \gg n$ . . . 133
3.4.3 Connectivity . . . 134
3.4.3.1 Class of Graphs of Euler Genus $g(n) = o\left(\frac{n}{\log^3(n)}\right)$ . . . 135
3.4.3.2 Class of Graphs of Euler Genus $O\left(\frac{n}{\log(n)}\right)$ . . . 135
3.4.3.3 Class of Graphs of Euler Genus $g(n) = \omega(n)$ . . . 136
3.4.4 Maximum Degree . . . 138
3.5 Discussion of Threshold Behaviour . . . 139
3.6 Unlabeled Graphs . . . 141
3.6.1 Lower Bounds . . . 142
3.6.2 Upper Bounds . . . 143
3.6.2.1 Orientable Case . . . 143
3.6.2.2 Non-orientable Case . . . 145
3.7 Further Open Problems and Considerations . . . 145

3.7.1 Further Consideration - The Class of Graphs of Hereditarily Bounded Genus 145
3.7.1.1 Number of Graphs of Hereditarily Bounded Genus . . . 145
3.7.1.2 Pendant Vertices 148
3.7.2 Further Open Problems 149

Bibliography **153**

List of Figures

2.1 The only possible types of K_4 in Γ_2 30
2.2 The only possible types of K_4 in Γ_3 31
2.3 These types of K_4 are not possible in Γ_3 32

3.1 $G \in \mathscr{A}_n$ to $G' \in \mathscr{A}_{(1+\delta)n}$ 74
3.2 $G \in \mathscr{A}_n$ to $G' \in \mathscr{A}_n$ 76
3.3 c_1 is surface-separating and contractible, c_2 is surface-separating, and c_3 and c_4 are non-surface-separating 78
3.4 $G \in \mathscr{A}_n^k$ to $G' \in \mathscr{A}_{n+1}^k$ 89
3.5 $G \in \mathscr{A}_n^k$ to $G' \in \mathscr{A}_{n+1}^{k+1}$ 90
3.6 $G \in \mathscr{A}_n^{k(n)}$ to $G' \in \mathscr{A}_n^{k(n)+2}$ 91
3.7 Wheel around vertex v 93
3.8 F_1 and F_2 incident to v 93
3.9 Moving anticlockwise starting at F_1 94
3.10 Splitting a vertex v of degree greater than three 108
3.11 The graph G' 123
3.12 $G \in \mathscr{B}_n$ to $G' \in \mathscr{A}_n$ 134
3.13 $G \in \mathscr{C}_n$ to $G' \in \mathscr{B}_n$ 137

Chapter 1

Introduction

1.1 Background

This thesis is split into two chapters, which each treat a different problem in random graph theory. While random graph theory is one of the youngest branches in Graph Theory, it has already been found to be of major importance beyond the field of Graph Theory. Erdős was the first to use a probabilistic construction to prove the existence of a graph with large girth and large chromatic number [6]. He continued to use probabilistic methods in Graph Theory to prove the existence of other graphs with seemingly contradictory properties. It was not until later that Erdős and Rényi became interested and started investigating random graphs in their own right. They introduced the $G(n, m)$ random graph model, where each graph is chosen uniformly at random from the set of all graphs on n vertices and m edges [7]. Gilbert introduced the so called Erdős–Rényi random graph model $G(n, p)$, which is the random graph on the vertex set $[n] = \{1, \cdots, n\}$ where each of the $\binom{n}{2}$ edges is present with probability p independently of the presence of any other edges [13]. The main ground work in the study of random graphs was however laid by Erdős and Rényi. They published a series of papers which laid the foundations of random graph theory and proved many fundamental results [8, 9, 10, 11]. In the eighties, the study of random graphs started to get much more attention and Béla Bollobás published the now considered classic book called Random Graphs [4].

The main objective of the study of Random Graphs tends to be to find which properties a graph from a given random graph model "typically" has. This means that we try to find which properties a random graph from the model possess with high probability, where with high probability means with probability tending to one as n tends to infinity, where n usually denotes the number of vertices. Further, we wish to obtain the distribution of the appearance of such properties in random graphs from a specified random graph model. In this thesis, we deal with two different random graph models.

The first chapter will treat a distributional problem in the Erdős–Rényi random graph model. In particular, we tackle the problem of finding a local central limit theorem for the number of K_4 subgraphs in the Erdős–Rényi random graph $G(n, p)$. The second chapter is devoted to a less classic random graph model. We investigate the random graph embeddable on a non-constant surface. This means more precisely that we consider graphs chosen uniformly at random from a class, where the class is defined to be the set of all graphs satisfying a certain criterion. In our case, the criterion is the embeddability of the graph on an orientable or non-orientable surface of Euler genus at most $g(n)$. To obtain results on the "typical" or the distribution of properties of such graphs, we must first obtain results on the sizes of such graph classes.

We are keeping further introductory material and discussion for each problem within the chapter concerning it.

1.2 Main Ideas in Thesis

In this section we will give a brief overview of the main proof ideas used in the following chapters. This chapter will often not be entirely rigorous as its purpose is mainly to give a better understanding of the ideas used in this thesis. For details, look in the chapters themselves.

1.2.1 Local Limit Theorem for Number of K_4 Subgraphs in $G(n,p)$

In the first chapter we use the fact that in order to prove a local limit theorem for the number of K_4 subgraphs in $G(n,p)$, where p is a fixed constant, it suffices to show that the characteristic function for the appearance of K_4 is close in L_1-distance to the characteristic function of a Gaussian distribution. It turns out that to show this it suffices to bound the absolute value of the characteristic function of the number of K_4 subgraphs sufficiently well above.

We do this by using different conditioning arguments for different ranges of t, where the characteristic function is a function in t. We usually condition on some set of edges. The set of edges is chosen in such a way, that they have some nice property with relatively high probability. We further show that as long as they possess this nice property, the conditional expectation and hence the characteristic function is small in absolute value. An outline of the main conditioning argument can be found in Section 2.6.

In our proofs we mostly use two different approaches. The first is to just condition on some set of edges in the graph. The other way is to split the vertices into two parts. We then condition on the edges inside one part. Instead of looking at the number of K_4 dependent on edges between the two parts, it suffices to look at the symmetric difference in numbers of K_4 between two independent copies of the edges between the two parts. Thus, by showing that the two independent copies of edges between the two parts produce a sufficiently similar number of K_4 subgraphs, we can show that the characteristic function is sufficiently small. For the arithmetic explanation of this, see Lemma 18.

In each of the different ranges of t, we make use of a combination of the conditioning arguments we just outlined. In the first two ranges, we simply condition on everything but some number of independent copies of a complete bipartite graph. Once we have conditioned on these, the actual number of K_4 subgraphs is the sum of a binomial random variable and the random variable counting the number of K_4 subgraphs with more than one edge in the independent copies of a complete bipartite graph. To bound the absolute value of the characteristic function above it thus suffices to bound the conditional expectation of the absolute value of the characteristic function of a Binomial random variable

above and show that the other variable is close to its expectation using concentration bounds. For the actual proofs of this, see Sections 2.7.2 and 2.7.3. In the next two ranges, we split the vertices into two parts, where both parts are relatively large, and then use a similar conditioning argument to the previous sections applied to the edges inside the second part of vertices. For the actual proofs of this, see Sections 2.7.4 and 2.7.5. In the final two ranges, we use the splitting vertices into two parts argument twice. In the first round, we split off a single vertex, u say. We now apply the splitting vertices step again on the graph excluding u and all edges incident to u. We pick the first part of vertices, V say, to be a vertex or a set of vertices that are connected to u in both independent copies of the edges between u and the rest of the graph. We can show that such a vertex or set of vertices exists in most cases. By applying the splitting the vertices argument to the set of vertices V and all vertices except u and V, we obtain a sufficiently good bound on the absolute value of the characteristic function. For the actual proofs of this, see Sections 2.7.6 and 2.7.7.

1.2.2 Graphs on Non-constant Surfaces

In the second chapter, we look at the classes of graphs embeddable on a surface of genus $g(n)$, where n is the number of vertices. Note that as soon as $g(n)$ is a non-constant function, generating function methods, which have been very powerful in the constant genus case, appear to break down.

In the first part of the chapter, we aim to find upper and lower bounds on the sizes of the class of graphs embeddable on a surface of Euler genus at most $g(n)$ for different functions $g(n)$. For the lower bounds, we use the fact that Euler's formula for cellularly embedded connected graphs on surfaces gives a lower bound on the number of edges in a connected graph cellularly embeddable in a surface of genus $g(n)$. We further use double counting arguments to relate graphs to graphs of lower genus and eventually planar graphs. This is useful since an asymptotic counting formula for the number of planar graphs on n vertices exists. For the exact results and proofs, see Section 3.3.1.

We now go on to consider upper bounds on the sizes of these graph classes. To find an upper bound, we show that any connected, cellular graph embedding on a surface can be constructed from some unicellular map on the same surface by labelling the vertices and then adding edges in the unique face of the embedding. In the orientable case an exact counting formula for the number of unicellular maps on n vertices exists. We can thus use this to find an upper bound on the number of connected graphs cellularly embeddable on a surface. The intuition behind this is that the unicellular map in some sense takes care of wrapping around the "complex" part of the surface structure and then all that is left to take care of is to fill the unique planar like face with edges to construct graphs

with multiple faces. In the non-orientable case the problem becomes more involved as no exact counting formula exists for the number of unicellular maps on a surface. We find an upper bound on the number of such maps by constructing any possible unicellular map by first contracting edges in a pre-cubic unicellular map and then adding vertices of degree two in the middle of existing edges. This is sufficient as an upper bound on the number of pre-cubic unicellular maps exists. For the details of these constructions, see Section 3.3.2.

The next part of the second chapter concerns proving the existence of a growth constant equal to the planar graph growth constant up to genus $g(n) = o\left(\frac{n}{\log^3(n)}\right)$. For this proof, we make use of the fact that for any graph embeddable on a surface an edge planarising set including no more than $O(\sqrt{ng\Delta})$ edges exists, where n is the number of vertices in the graph, g is the genus of the surface it is embedded on and Δ is the maximum degree in the graph. Generally, no useful upper bound on the maximum degree in a graph embedded on a surface exists. However, we make use of a proof similar to Colin McDiarmid's proof of an upper bound on the maximum degree in a planar graph (see Theorem 60), to show that as long as a significant fraction of graphs in the graph class have linearly many pendant vertices, a significant fraction of graphs in the graph class has maximum degree of order at most $O(\log(n))$. We can further show, using a proof similar to the proof of the Pendant Appearances Theorem for planar graphs (see Theorem 59), that as long as the size of the graph class does not grow too fast in n, linearly many pendant vertices exist with large probability. We make use of the upper bounds found in the previous section to show that the graph classes cannot grow too quickly too often. In other words, there are an infinite number of n such that the growth in the size of the graph class when moving from n to $n+1$ vertices is sufficiently small to show that a large portion of graphs in the class has linearly many pendant vertices. By knowing that these n occur sufficiently often, we can show that the limsup of the "growth rate" is equal to the planar graph growth constant. Equally, since all planar graphs are contained in the class, the liminf of the "growth rate" is also equal to the planar graph growth constant. This gives us the desired result. For details on this proof, see Section 3.3.2.3.

We also give an alternative proof for the existence of a growth constant up to genus $g(n) = o\left(\frac{n}{\log^{3+\epsilon}(n)}\right)$ for any $\epsilon > 0$. We do this by first proving the existence of a growth constant up to genus $g(n) = o\left(\frac{n}{\log^4(n)}\right)$ by repeatedly splitting the vertices along a non-surface separating cycle meeting the graph only at vertices. In each step, we increase the number of vertices but also decrease the genus. By using a double counting argument and since we know an asymptotic counting formula for the number of planar graphs, we find a bound on the number of graphs embeddable on the surface of genus at most $g(n)$. It turns out that this bound is sufficiently good to show that up to genus $g(n) = o\left(\frac{n}{\log^4(n)}\right)$

a growth constant equal to the planar graph growth constant exists. In the next step, we try to use this knowledge to show the existence of a growth constant beyond $o\left(\frac{n}{\log^4(n)}\right)$. We thus construct from a graph with relatively large genus (call this type of graph type A) a graph with relatively smaller genus (call this type of graph type B) by adding to the graph a big connected planar component (call this type of graph type C), thus bringing the overall genus relative to the number of vertices down. As long as the genus of the resulting graph is sufficiently small that we know that a growth constant exists, we have a pretty good upper bound on the number of such graphs. We can thus build a double counting argument. Since we also know very good bounds on the number of planar graphs, we can, by using the relationship that

$$\text{number of graphs of type } A \cdot \text{number of graphs of type } B \leqslant \text{number of graphs of type } C \ ,$$

which we know from our construction, find a bound on the number of graphs of type A. By choosing the sizes of these graphs correctly we can thus repeatedly get a little further. By using a supremum argument, we can prove the existence of a growth constant up to genus at most $o\left(\frac{n}{\log^{3+\epsilon}(n)}\right)$ this way.

In the next section, we look at common properties of graphs in these graph classes. Most of the results are proven by double counting arguments. We use a combination of Euler's formula, the fact that we know all bridge-addable classes have with high probability few components and the fact that as soon as the genus satisfies $g(n) = \omega(n)$ we have with high probability $\omega(n)$ edges present in the graph.

We then extend some of our proofs to also work in the unlabelled case (see Section 3.6) before moving on to introduce the notion of hereditary bounded genus. A graph has orientable hereditary genus $g(n)$ if it is embeddable in the orientable surface of Euler genus $g(n)$ and such that for any set of m vertices V, the induced graph $G[V]$ is embeddable in an orientable surface of Euler genus at most $g(m)$. To prove an upper bound on the number of such graphs we show that any graph constructed by identifying the vertices of $k+1$ planar graphs has orientable hereditary genus at most $2kn$. For the full details of this see Section 3.7.1.

Chapter 2

Local Limit Theorem for Number of K_4 Subgraphs in $G(n,p)$

We consider the number of subgraphs isomorphic to K_4, the complete graph on four vertices, in the Erdős–Rényi random graph $G(n,p)$. Recall that $G(n,p)$ is the graph on n vertices which has each of the $\binom{n}{2}$ possible edges present with probability p independently of the presence of any other edges. We denote by $\mathbf{S}_n$ the number of K_4 subgraphs in $G(n,p)$. Note that $\mu_n = \mathbb{E}[\mathbf{S}_n] = p^6\binom{n}{4}$ and $\sigma_n^2 = \mathrm{Var}(\mathbf{S}_n) = \Theta\left(n^6\right)$. Our main result is a local limit theorem for $\mathbf{S}_n$ for a fixed edge probability p. The result says that if p is constant, then the distribution of $\mathbf{S}_n$ is point-wise approximated by a discrete Gaussian distribution, so

$$\mathbb{P}\left(\mathbf{S}_n = k\right) = \frac{1}{\sigma_n\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{k-\mu_n}{\sigma_n}\right)^2} + o\left(n^{-3}\right) \ . \tag{2.1}$$

In other words, for every $k = \mu_n \pm O\left(n^3\right)$ we find the probability that $G(n,p)$ has exactly k subgraphs that are isomorphic to K_4 up to a $(1+o(1))$ factor.

2.1 Introduction

In this chapter, we will be considering the Erdős–Rényi random graph $G(n,p)$ and in particular the distribution of the number of subgraphs isomorphic to the complete graph on four vertices K_4. The Erdős–Rényi random graph $G(n,p)$ is defined as follows:

Definition. *Given an integer $n \geqslant 1$ and a real number $p \in [0,1]$, we define the **Erdős–Rényi random graph** $G(n,p)$ to be the graph on the vertex set $[n] = \{1,\ldots,n\}$ such that each of the $\binom{n}{2}$ edges is present with probability p independently of the presence of any other edges.*

By this definition, for any graph H on n vertices with e edges,

$$\mathbb{P}(G(n,p) = H) = p^e(1-p)^{\binom{n}{2}-e} \ .$$

And further $G(n, 1/2)$ is hence the graph chosen uniformly at random from the class of all graphs on $[n]$. It is worth noting that here and for the remainder of this chapter we will be working exclusively with labelled graphs.

A classical question in the study of random graphs concerns the probability of existence and distribution of the number of subgraphs isomorphic to some fixed graph H in $G(n, p)$. In this chapter we consider the distribution of the number of subgraphs isomorphic to K_4, the complete graph on four vertices, in the Erdős–Rényi random graph $G(n, p)$ for constant edge probability p. Let $\mathbf{S}_n$ denote the number of K_4 subgraphs in $G(n, p)$. Note that $\mu_n = \mathbb{E}[\mathbf{S}_n] = p^6\binom{n}{4}$ and $\sigma_n^2 = \mathrm{Var}(\mathbf{S}_n) = \Theta\left(n^6\right)$ (see Lemma 19). Our main result in this chapter is a local limit theorem for $\mathbf{S}_n$ for a fixed edge probability p. The result says that if p is constant, then the distribution of $\mathbf{S}_n$ is point-wise approximated by a discrete Gaussian distribution, so

$$\mathbb{P}\left(\mathbf{S}_n = k\right) = \frac{1}{\sigma_n\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{k-\mu_n}{\sigma_n}\right)^2} + o\left(n^{-3}\right) \ . \tag{2.2}$$

In other words, for every $k = \mu_n \pm O\left(n^3\right)$ we find the probability that $G(n, p)$ has exactly k subgraphs that are isomorphic to K_4 up to a $(1 + o(1))$ factor. We will now give a brief overview of the history of relevant results in subgraph counts before continuing with our results.

2.1.1 Graph Containment Thresholds

The subgraph containment problem actually originated with a result by Erdős and Rényi, which established a sharp threshold for the containment of any so called balanced subgraph H in the random graph $G(n, p)$, where a balanced graph is defined as follows:

Definition. *For any graph G on $|G|$ vertices and with $e(G)$ edges, we define its **density** $d(G)$ and **maximum subgraph density** $m(G)$ to be*

$$d(G) := \frac{e(G)}{|G|}$$
$$m(G) := \max\{d(H) \mid H \subset G\} \ .$$

*A graph G is then called **balanced** if $d(G) = m(G)$. Furthermore, G is **strictly balanced** if for any subgraph H of G we have $d(H) = m(G)$ if and only if $H = G$.*

Here and in the remainder of this chapter, the number of vertices of the graph G is denoted by $|G|$, the number of edges of the graph G is denoted by $e(G)$ and we write $H \subset G$ if H is a subgraph of G. We can now state the threshold result by Erdős and Rényi for balanced graphs.

Theorem 1. *[8] If G is balanced then*

$$\lim_{n\to\infty} \mathbb{P}(G \subset G(n,p)) = \begin{cases} 0 & \text{if } np^{d(G)} = o(1), \\ 1 & \text{if } np^{d(G)} \to \infty. \end{cases} \tag{2.3}$$

The proof of this result is based on the method of moments. It then took twenty-one years until this result was generalised to the containment of any subgraph G by Bollobás, whose proof is actually based on Theorem 1. By now, several different proofs of this theorem are known.

Theorem 2. *[3]* ***(Threshold Theorem)***
For any graph G,

$$\lim_{n\to\infty} \mathbb{P}(G \subset G(n,p)) = \begin{cases} 0 & \text{if } np^{m(G)} = o(1), \\ 1 & \text{if } np^{m(G)} \to \infty. \end{cases} \tag{2.4}$$

A few years later, exponential bounds on the probability of the nonexistence of copies of any graph G in $G(n,p)$ were found [16]. While these are interesting in their own right, they also directly imply the Threshold Theorem (see Theorem 2). The threshold behaviour of many different properties of the Erdős–Rényi random graph have been studied in much detail by now. However, the answer to the containment problem of subgraphs in the random graph $G(n,p)$ gave rise to further questions, like the distribution of the number of subgraphs isomorphic to a given graph G on and beyond the threshold.

2.1.2 Central Limit Theorems

Since Bollobás basically gave a full solution to the containment problem, several results about the distribution of the number of these subgraphs on and beyond the threshold have been found. For some fixed graph G, we denote by $X_n(G)$ the random variable equal to the number of copies of G in $G(n,p)$ and denote by

$$\tilde{X}_n(G) = \frac{X_n(G) - \mathbb{E}(X_n(G))}{\sqrt{\mathrm{Var}(X_n(G))}}$$

the normalised number of copies of G in $G(n,p)$. At the threshold, the distributional question has been partially solved.

Theorem 3. *[3][17] Let G be any fixed graph. Assume that $np^{d(G)} \sim c$ for some constant $c > 0$. Then $X_n(G)$ converges in distribution to a Poisson random variable with parameter $\mathbb{E}[X_n(G)] = c^{|G|}/\mathrm{aut}(G)$.*

Beyond the threshold, one would expect the normalised number of subgraphs isomorphic to G to roughly follow a normal distribution. This is the case since one would expect its distribution close to the threshold to be similar to its distribution at the threshold. It has been shown that the number of subgraphs isomorphic to G is close to a Poisson distribution with parameter $\mathbb{E}[X_n(G)]$ at the threshold. We assume this is still roughly the case just beyond the threshold, so the normalised number of subgraphs isomorphic to G roughly follows a Poisson distribution with parameter $\mathbb{E}[X_n(G)]$ still. Since however $\mathbb{E}[X_n(G)] \to \infty$ beyond the threshold, this implies that the normalised number of subgraphs isomorphic to G behaves close to a Normal random variable.

The first general result in the direction of a distributional result beyond the threshold was proven by Karoński and Ruciński [17] who proved the asymptotic normality of the normalised subgraph count of G close to the threshold. In particular, they proved that $\tilde{X}_n(G)$ converges in distribution to a standard normal random variable under the restriction that G must be strictly balanced and for every $\epsilon > 0$, we have $np^{d(G)} = o(n^{\epsilon})$. The method used for the proof is the moment method. The moment method relies on the fact that the convergence of the moments of a sequence of random variables to the moments of the distribution implies convergence in distribution.

Moving further away from the threshold, for p constant, which is also the problem we will be considering more closely in this chapter of the thesis, Nowicki [18] was the first to prove the convergence in distribution of $\tilde{X}_n(G)$ to a standard normal distribution. His method of proof uses a technique from asymptotic theory of statistics, known as incomplete U-statistics. In [20], Ruciński fully determined when $\tilde{X}_n(G)$ converges in distribution to a standard normal, which is proved by considering the moments of $X_n(G)$.

Theorem 4. *[20] For an arbitrary graph G, $\tilde{X}_n(G)$ converges in distribution to a standard normal distribution if and only if $np^{m(G)} \to \infty$ and $n^2(1-p) \to \infty$.*

By making extensive use of Stein's normal approximation, strong quantitative bounds on the rate of convergence to the normal distribution were found by Barbour, Karoński and Ruciński[1]. Three main methods of proof have been used in the proof of central limit theorems for the subgraph count so far: applications of the moment method, Stein's method or the U-statistics. However, none of these techniques appear to be strong enough to attain point probability mass estimates for constant edge probability values p. Nonetheless, all of these results suggest the existence of a local central limit theorem for the number of copies of any fixed graph G in $G(n,p)$ for p constant. In this chapter, we will prove such a theorem for $G = K_4$, the complete graph on four vertices.

2.1.3 Local Limit Theorem for Triangle Count

Recently, Gilmer and Kopparty managed to prove a local central limit theorem for the triangle count in $G(n,p)$ for p constant. We denote by $\tilde{S}_n$ the random variable expressing the number of triangles present in $G(n,p)$ and by $\tilde{R}_n$ the normalised number of triangles present in $G(n,p)$, so $\tilde{R}_n = \frac{S_n - p^3\binom{n}{3}}{\sigma_n}$.

Theorem 5. *[14]* ***(Local limit law for triangles in*** $G(n,p)$***)***
Let

$$p_n(x) = \mathbb{P}(\tilde{R}_n = x) \text{ for } x \in \mathbb{L}_n = \left\{ \frac{k - p^3\binom{n}{3}}{\sigma_n} \mid k \in \mathbb{Z} \right\}$$

and

$$\mathcal{N}(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \text{ for } x \in (-\infty, \infty).$$

Then as $n \to \infty$*,*

$$\sup_{x \in \mathbb{L}_n} |\sigma_n p_n(x) - \mathcal{N}(x)| \to 0.$$

Equivalently, we have that for all n, for all $k \in \mathbb{Z}$,

$$\mathbb{P}(\tilde{S}_n = k) = \frac{1}{\sigma_n} \cdot \mathcal{N}\left(\frac{k - p^3\binom{n}{3}}{\sigma_n} \right) + o\left(\frac{1}{n^2} \right) ,$$

where the $o(1)$ term goes to 0 as $n \to \infty$ uniformly in k.

Their proof is based on bounding the characteristic function of the triangle count in $G(n,p)$ by using different conditioning arguments. In our proof, we will be using a similar proof technique for proving a local central limit theorem for the number of K_4.

2.2 Statement of Main Result

We now state our main theorem, which we will prove in the remaining sections.

Theorem 6. ***(Local limit law for number of*** $\mathbf{K_4}$ ***in*** $\mathbf{G(n,p)}$***)***
Let $p \in (0,1)$ *be a fixed constant. Denote by* $\mathbf{S}_n$ *the number of* K_4 *subgraphs in* $G(n,p)$*. For all* $k = k(n) \in \mathbb{Z}$*, the random variable* $\mathbf{S}_n$ *satisfies the following local limit theorem*

$$\mathbb{P}\left(\mathbf{S}_n = k\right) = \frac{1}{\sigma_n} \cdot \phi\left(\frac{k - \mu_n}{\sigma_n} \right) + o\left(n^{-3}\right) , \tag{2.5}$$

where the $o(1)$ *term goes to* 0 *as* $n \to \infty$ *uniformly in* k*,* $\mu_n = \mathbb{E}[\mathbf{S}_n] = p^6\binom{n}{4}$ *is the expectation of* $\mathbf{S}_n$*,* $\sigma_n^2 = \mathrm{Var}(\mathbf{S}_n) = \Theta(n^6)$ *is the variance of* $\mathbf{S}_n$ *and* $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ *is the probability density function of a standard normal distribution.*

(Remark: since the draft of this chapter was written, Berkowitz has posted a preprint on arXiv giving a more general version of this result, applying to K_r for every fixed r, see [2].)

2.3 Overview of Proof Method

The high level structure of our proof follows that of the local limit theorem for the number of K_3 in $G(n,p)$ by Justin Gilmer and Swastik Kopparty in [14]. To show that the distribution of $\frac{\mathbf{S}_n-\mu_n}{\sigma_n}$ is close point-wise to a standard normal distribution, it suffices to show that their characteristic functions (Fourier transforms) are close in L_1 distance. The characteristic function ϕ of a random variable $\mathbf{X}$ is defined as

$$\phi(t) = \mathbb{E}\left[e^{it\mathbf{X}}\right] \ .$$

Concretely, writing $\psi_n(t) = \mathbb{E}\left[e^{it\mathbf{S}_n}\right]$ for the characteristic function of the number of K_4 in $G(n,p)$ and $\psi_n^*(t) = \mathbb{E}\left[e^{it(\mathbf{S}_n-\mu_n)/\sigma_n}\right]$ for the normalised form of ψ_n it suffices to show that

$$\int_{-\pi\sigma_n}^{\pi\sigma_n} \left|\psi_n^*(t) - e^{-t^2/2}\right| dt = o(1) \ . \tag{2.6}$$

We give more detail on why it is sufficient to prove this at the beginning of the next section. The central limit theorem for the number of K_4 can be used to bound the above integral in the range $(-A, A)$ for any constant A. We can furthermore bound

$$\int_{A<|t|<\pi\sigma_n} \left|e^{-t^2/2}\right| dt \tag{2.7}$$

by an arbitrarily small constant by choosing A large enough. This implies that for equation (2.6) to be valid, it suffices to show that for every $\epsilon > 0$ there exists a constant $A \in \mathbb{R}$ such that for n large enough

$$\int_{A<|t|<\pi\sigma_n} |\psi_n^*(t)|\, dt < \epsilon \ . \tag{2.8}$$

We will achieve this by bounding $|\psi_n^*(t)|$ by $t^{-(1+\delta)}$ for some fixed $\delta > 0$ by using conditioning arguments. Different conditioning arguments will be used for different ranges of t. In total, we will split $[A, \pi\sigma_n]$ into seven different ranges and use different arguments in each range.

Note that the Berkowitz paper uses the same fundamental strategy as the K_3 paper, which is also the strategy we use. It turns out that Berkowitz's key lemma is a more general version of the proof method we use in Section 2.7.7, but applied in a pretty different way.

2.4 Fundamental Proof Strategy

Let $\psi_n^*(t) = \mathbb{E}\left[e^{it(\mathbf{S}_n-\mu_n)/\sigma_n}\right]$ be the characteristic function of the normalised number of K_4 subgraphs in $G(n,p)$. As mentioned before, the general bounding argument for

the characteristic function follows that of [14], however the bounding arguments in the different regions of t will be different. For completeness, we give an overview of the argument here. The Fourier inversion formula for lattices (see [12, Theorem 4, Chapter 15.3]) gives us

$$\mathbb{P}\left(\frac{\mathbf{S}_n - \mu_n}{\sigma_n} = x\right) = \frac{1}{2\pi\sigma_n}\int_{-\pi\sigma_n}^{\pi\sigma_n} e^{-itx}\,\psi_n^*(t)dt \tag{2.9}$$

for x such that $x\sigma_n + \mu_n \in \mathbb{Z}$. The standard Fourier inversion formula (for $\mathbb{R}$), along with the well–known formula for the Fourier transform of the standard normal distribution ϕ gives us

$$\phi(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}e^{-t^2/2}dt\ . \tag{2.10}$$

Therefore,

$$\left|\sigma_n \cdot \mathbb{P}\left(\frac{\mathbf{S}_n - \mu_n}{\sigma_n} = x\right) - \phi(x)\right| \leqslant \frac{1}{2\pi}\int_{-\pi\sigma_n}^{\pi\sigma_n}\left|\psi_n^*(t) - e^{-t^2/2}\right|dt + \frac{1}{\pi}\int_{\pi\sigma_n}^{\infty} e^{-t^2/2}dt\ .$$

Since the second term goes to zero as n goes to infinity, to prove Theorem 6 it suffices to show that

$$\int_{-\pi\sigma_n}^{\pi\sigma_n}\left|\psi_n^*(t) - e^{-t^2/2}\right|dt = 2\int_0^{\pi\sigma_n}\left|\psi_n^*(t) - e^{-t^2/2}\right|dt \tag{2.11}$$

tends to zero as n tends to infinity. Let $A > 0$ be a large constant, which we determine later. As shown previously, to bound $\int_0^{\pi\sigma_n}\left|\psi_n^*(t) - e^{-t^2/2}\right|dt$ above it suffices to show that for every $\epsilon > 0$ and some constant $A = A(\epsilon) > 0$ there exists an $N \in \mathbb{N}$ such that for every $n \geqslant N$

$$\int_A^{\pi\sigma_n}|\psi_n^*(t)|\,dt < \epsilon\ .$$

This is achieved by bounding $|\psi_n^*(t)|$ above by at least $t^{-(1+\delta)}$ for some $\delta > 0$. We now change the variable of integration to $\theta = t/\sigma_n$ as this will simplify notation later on. We thus further define $\widetilde{\psi}_n^*(\theta) := \psi_n^*(\sigma_n\theta)$. To bound $\widetilde{\psi}_n^*$, we divide the region of integration into the following seven different regions for θ

- $R_1 = (0, A/\sigma_n)$
- $R_2 = (A/\sigma_n, n^{-2.45})$
- $R_3 = (n^{-2.45}, n^{-2.1})$
- $R_4 = (n^{-2.1}, n^{-1.96})$
- $R_5 = (n^{-1.96}, n^{-1.4})$
- $R_6 = (n^{-1.4}, n^{-0.6})$
- $R_7 = (n^{-0.6}, \pi)$.

We will give different arguments for bounding the characteristic function for the normalised number of K_4 in $G(n,p)$ in these seven regions.

Lemma 7. *Let A be a fixed positive real number. Then*

$$\int_0^{A/\sigma_n} \sigma_n \left| \widetilde{\psi}_n^*(\theta) - e^{-(\sigma\theta)^2/2} \right| dt \to 0 \ , \tag{2.12}$$

as $n \to \infty$.

Lemma 8. *There exists a constant $D_2 = D_2(p)$ such that for all $\theta \in R_1 \cup R_2$ we have*

$$\left| \widetilde{\psi}_n^*(\theta) \right| \leqslant D_2 \, (\sigma\theta)^{-1.5} \ . \tag{2.13}$$

Lemma 9. *There exists a constant $D_3 = D_3(p)$ such that for all $\theta \in R_3$ we have*

$$\left| \widetilde{\psi}_n^*(\theta) \right| \leqslant D_3 \, (\sigma\theta)^{-1.45} \ . \tag{2.14}$$

Lemma 10. *There exists a constant $D_4 = D_4(p)$ such that for all $\theta \in R_4$ we have*

$$\left| \widetilde{\psi}_n^*(\theta) \right| \leqslant D_4 \, e^{-n^{0.02}} \ . \tag{2.15}$$

Lemma 11. *There exists a constant $D_5 = D_5(p)$ such that for all $\theta \in R_5$ we have*

$$\left| \widetilde{\psi}_n^*(\theta) \right| \leqslant D_3 \, e^{-n^{0.06}} \ . \tag{2.16}$$

Lemma 12. *There exists a constant $D_6 = D_6(p)$ such that for all $\theta \in R_6$ we have*

$$\left| \widetilde{\psi}_n^*(\theta) \right| \leqslant D_6 \, (\sigma\theta)^{-25} \ . \tag{2.17}$$

Lemma 13. *There exists a constant $D_7 = D_7(p)$ such that for all $\theta \in R_7$ we have*

$$\left| \widetilde{\psi}_n^*(\theta) \right| \leqslant D_7 \, e^{-n^{0.8}} \ . \tag{2.18}$$

The proofs of these lemmas will follow in the remaining subsections. We will now show how these lemmas can be used to bound the integral in equation (2.11). The change of variable of integration from t to θ will aid to make arguments more clear later on. For the rest of this section it is however more natural to use the variable t, so we will revert back to using t for the remainder of this section. Further let $R = R_2 \cup R_3 \cup R_4 \cup R_5 \cup R_6 \cup R_7$. Note that $\widetilde{\psi}^*(\theta)$ being small on R implies $\psi^*(t)$ being small on $[A, \sigma_n \pi] = \sigma_n R$, so from the bounds on θ in the above lemmas follow sufficiently good bounds on t. By Lemma 7, we have for any fixed constant A

$$\int_{\sigma_n R_1} \left| \psi_n^*(t) - e^{-t^2/2} \right| dt \to 0 \ . \tag{2.19}$$

We thus have the following

$$\int_{\sigma_n R} \left|\psi_n^*(t) - e^{-t^2/2}\right| dt \leqslant \int_{\sigma_n R} |\psi_n^*(t)|\, dt + \int_{\sigma_n R} \left|e^{-t^2/2}\right| dt \ . \tag{2.20}$$

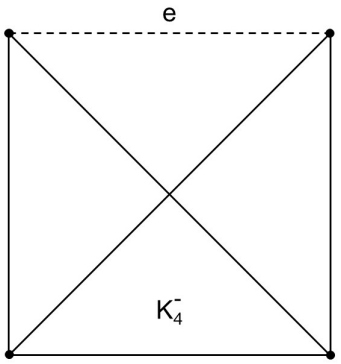
Combining Lemmas 8 - 13 we have that there exists a constant $D = D(p)$ and a $\delta > 0$ such that $|\psi_n^*(t)| \leqslant D \cdot t^{-(1+\delta)}$ for all n and all t with $t \in (1, \pi\sigma_n]$. Therefore,

$$\int_{\sigma_n R} \left|\psi_n^*(t) - e^{-t^2/2}\right| dt \leqslant \int_{\sigma_n R} \left|\frac{D}{t^{1+\delta}}\right| dt + \int_{\sigma_n R} \left|e^{-t^2/2}\right| dt \ . \tag{2.21}$$

Then as both $Dt^{-(1+\delta)}$ and $e^{-t^2/2}$ have finite integral over the range $(1, \infty)$, equation (2.11) can be bound above by ϵ for any $\epsilon > 0$ by choosing $A = A(\epsilon, p)$ large enough. To prove Theorem 6 it is hence sufficient to prove Lemmas 7 - 13.

2.5 Notation and Preliminaries

Recall that the Erdős–Rényi random graph, denoted by $G(n, p)$, is the random graph on n vertices with each possible edge appearing independently with probability p. In what follows we will always take p to be a fixed constant. For a random variable $\mathbf{X}$ and a set of edges $A \subseteq E(K_n)$, so a set of possible edges in $G(n, p)$, we will denote by $\mathbb{E}_A[\mathbf{X}]$ the expectation averaging only over the edges inside A and not any other edges. Thus $\mathbb{E}_A[\mathbf{X}]$ is itself a random variable depending only on the edges outside A. Thus $\mathbb{E}[\mathbf{X}] = \mathbb{E}_{E(K_n)}[\mathbf{X}]$. For an edge e, we define a K_4^- *based at* e to be a subgraph H such that $H \cup \{e\}$ is a copy of K_4.



In what follows, we will frequently use the following lemmas. We let $||x||$ denote the distance from x to the nearest integer.

Lemma 14. *[14, Lemma 1] Let B be a Bernoulli random variable with parameter p. Then*

$$\left|\mathbb{E}[e^{i\gamma B}]\right| \leqslant 1 - 8p(1-p)\, \left|\left|\frac{\gamma}{2\pi}\right|\right|^2 \ . \tag{2.22}$$

Lemma 15. *Let* $\mathbf{X}$ *be a random variable with mean zero. Then*

$$\mathbb{E}[\left|e^{i\gamma\mathbf{X}}-1\right|] \leqslant \gamma \cdot \sqrt{\mathrm{Var}(\mathbf{X})}\ . \tag{2.23}$$

Proof. For any real number α, we have $|e^{i\alpha}-1| \leqslant 2\pi \left|\left|\frac{\alpha}{2\pi}\right|\right|$, and for any random variable $\mathbf{X}$, we have $\mathbb{E}\left[|\mathbf{X}|\right]^2 \leqslant \mathbb{E}\left[\mathbf{X}^2\right]$. Thus

$$\begin{aligned}
\mathbb{E}\left[\left|e^{i\gamma\mathbf{X}}-1\right|\right] &\leqslant 2\pi \cdot \mathbb{E}\left[\left|\left|\frac{\gamma\,\mathbf{X}}{2\pi}\right|\right|\right] \\
&\leqslant 2\pi \cdot \mathbb{E}\left[\left|\frac{\gamma\,\mathbf{X}}{2\pi}\right|\right] \\
&\leqslant 2\pi \cdot \sqrt{\mathbb{E}\left[\left(\frac{\gamma\,\mathbf{X}}{2\pi}\right)^2\right]} \\
&= 2\pi \cdot \sqrt{\mathrm{Var}\left(\frac{\gamma\,\mathbf{X}}{2\pi}\right)} \quad \text{since } \mathbf{X} \text{ has mean zero} \\
&= \gamma \cdot \sqrt{\mathrm{Var}\left(\mathbf{X}\right)}\ ,
\end{aligned}$$

as required. ■

Lemma 16. *Let* $A_1,\ldots,A_k$ *be random variables and let* $\alpha_i = \mathbb{E}[A_i]$ *be their expectations. Then*

$$\left|\mathbb{E}\left[e^{i\theta\left(\sum_{i=1}^k A_i\right)}\right]\right| \leq \left|\mathbb{E}\left[e^{i\theta A_1}\right]\right| + \sum_{i=2}^{k} \mathbb{E}\left[\left|e^{i\theta(A_i-\alpha_i)}-1\right|\right]\ . \tag{2.24}$$

Proof. Assume without loss of generality that the mean of A_i is equal to zero. Then

$$\begin{aligned}
\left|\mathbb{E}\left[e^{i\theta\left(\sum_{i=1}^k A_i\right)}\right]\right| &= \left|\mathbb{E}\left[e^{i\theta A_1} + \sum_{j=2}^{k}\left(e^{i\theta\left(\sum_{l=1}^{j} A_l\right)} - e^{i\theta\left(\sum_{l=1}^{j-1} A_l\right)}\right)\right]\right| \\
&\leqslant \left|\mathbb{E}\left[e^{i\theta(A_1)}\right]\right| + \sum_{j=2}^{k} \mathbb{E}\left[\left|\left(e^{i\theta\left(\sum_{l=1}^{j} A_l\right)} - e^{i\theta\left(\sum_{l=1}^{j-1} A_l\right)}\right)\right|\right] \\
&= \left|\mathbb{E}\left[e^{i\theta A_1}\right]\right| + \sum_{j=2}^{k} \mathbb{E}\left[\left|e^{i\theta A_j}-1\right|\right]\ ,
\end{aligned} \tag{2.25}$$

as required. ■

Thus far, we have used slightly informal notation when denoting that a random variable depends on a set of edges, such as denoting by $\mathbf{X}(E)$ a random variable that depends only on the set of edges E. This is however an abuse of notation, since E is a set of edges in K_n and not a random variable. In certain places, we will need to consider two independent copies of the same set of edges. For this, it is necessary to distinguish between the set of possible edges E and the random variable representing which of these edges are

present. We thus denote by X_E the random vector denoting which of the edges inside E are present and by x_E one fixed observation of the random vector X_E. In the following two lemmas, we will be considering two independent copies of the indicator random variable for the edges inside a set A. Formally, these two independent copies will be denoted by X_A and X'_A. When using the shorthand notation, we will abuse notation and denote by A and A' the vectors X_A and X'_A.

Lemma 17. *[14] Let A and B be two disjoint subsets of the edges of K_n and let X_A and X_B be the random vectors denoting which of the edges inside A and B are present. Furthermore, let X'_A be an independent copy of X_A, let $\mathbf{X}(X_A, X_B)$ be a random variable depending on X_A and X_B and $\mathbf{Z}(X_B)$ a random variable depending only on B and not on A. Then*

$$\left|\mathbb{E}_{X_A,X_B}\left[e^{i\gamma(\mathbf{X}(X_A,X_B)+\mathbf{Z}(X_B))}\right]\right|^2 \leqslant \mathbb{E}_{X_A,X'_A,X_B}\left[e^{i\gamma\left(\mathbf{X}(X_A,X_B)-\mathbf{X}(X'_A,X_B)\right)}\right] \ . \tag{2.26}$$

Or in informal notation

$$\left|\mathbb{E}_{A,B}\left[e^{i\gamma(\mathbf{X}(A,B)+\mathbf{Z}(B))}\right]\right|^2 \leqslant \mathbb{E}_{A,A',B}\left[e^{i\gamma(\mathbf{X}(A,B)-\mathbf{X}(A',B))}\right] \ . \tag{2.27}$$

Proof. We have the following

$$\begin{aligned}
\left|\mathbb{E}_{X_A,X_B}\left[e^{i\gamma(\mathbf{X}(X_A,X_B)+\mathbf{Z}(X_B))]}\right]\right|^2 &\leqslant \mathbb{E}_{X_B}\left[\left|\mathbb{E}_{X_A}\left[e^{i\gamma(\mathbf{X}(X_A,X_B)+\mathbf{Z}(X_B))}\right]\right|^2\right] \\
&= \mathbb{E}_{X_B}\left[\left|e^{i\gamma\mathbf{Z}(X_B)}\cdot\mathbb{E}_{X_A}\left[e^{i\gamma\mathbf{X}(X_A,X_B)}\right]\right|^2\right] \\
&= \mathbb{E}_{X_B}\left[\left|\mathbb{E}_{X_A}\left[e^{i\gamma\mathbf{X}(X_A,X_B)}\right]\right|^2\right] \\
&= \mathbb{E}_{X_B}\left[\mathbb{E}_{X_A,X'_A}\left[e^{i\gamma\left(\mathbf{X}(X_A,X_B)-\mathbf{X}(X'_A,X_B)\right)}\right]\right] \ ,
\end{aligned}$$

as required. ∎

The following lemma is an extension of Lemma 17, which does not appear in [14], but is needed when considering K_4. As we will usually be using the shorter notation when applying these lemmas, we will state the following lemma in the informal notation for clarity of notation.

Lemma 18. *Let A, C and D be three disjoint sets of edges and let A' be an independent copy of A. Furthermore, let $\mathbf{V}(A, C, D)$ be a random variable depending on the sets A, C and D, and similarly let $\mathbf{W}(A, C)$, $\mathbf{X}(A, D)$, $\mathbf{Y}(C)$ and $\mathbf{Z}(D)$ be random variables. Let $\mathbf{S} = \mathbf{V}(A, C, D) + \mathbf{W}(A, C) + \mathbf{X}(A, D) + \mathbf{Y}(C) + \mathbf{Z}(D)$. Then*

$$\left|\mathbb{E}_{A,C,D}\left[e^{i\gamma\mathbf{S}}\right]\right|^2 \leqslant \mathbb{E}_{A,A',D}\left[\left|\mathbb{E}_{C}\left[e^{i\gamma(\mathbf{V}(A,C,D)-\mathbf{V}(A',C,D)+\mathbf{W}(A,C)-\mathbf{W}(A',C))}\right]\right|\right] \ .$$

Proof. Applying Lemma 17 with $B = C \cup D$, we see that

$$\begin{aligned}\left|\mathbb{E}_{A,C,D}\left[e^{i\gamma \mathbf{S}}\right]\right|^2 &\leqslant \mathbb{E}_{A,A'}\left[\mathbb{E}_{C,D}\left[e^{i\gamma(\mathbf{V}(A,C,D)-\mathbf{V}(A',C,D)+\mathbf{W}(A,C)-\mathbf{W}(A',C)+\mathbf{X}(A,D)-\mathbf{X}(A',D))}\right]\right] \\ &= \mathbb{E}_{A,A',D}\left[e^{i\gamma(\mathbf{X}(A,D)-\mathbf{X}(A',D))} \cdot \mathbb{E}_{C}\left[e^{i\gamma(\mathbf{V}(A,C,D)-\mathbf{V}(A',C,D)+\mathbf{W}(A,C)-\mathbf{W}(A',C))}\right]\right] \\ &\leqslant \mathbb{E}_{A,A',D}\left[\left|\mathbb{E}_{C}\left[e^{i\gamma(\mathbf{V}(A,C,D)-\mathbf{V}(A',C,D)+\mathbf{W}(A,C)-\mathbf{W}(A',C))}\right]\right|\right] \ .\end{aligned}$$

■

We now further show that the variance of the number of appearances of K_4 in $G(n,p)$ is of order $\Theta(n^6)$.

Lemma 19. *Let $S_n = \sum_C \mathbf{I}_C$ be the number of K_4 subgraphs in $G(n,p)$, where p is constant. Then the variance of S_n, $\mathrm{Var}(S_n)$, satisfies*

$$\mathrm{Var}(S_n) = \Theta\left(n^6\right) \ .$$

Proof. Expanding the variance of S_n, we obtain

$$\begin{aligned}\mathrm{Var}(S_n) &= \mathbb{E}[S_n^2] - \mathbb{E}[S_n]^2 \\ &= \sum_C \mathbb{E}[\mathbf{I}_C^2] - \sum_C \mathbb{E}[\mathbf{I}_C]^2 + \sum_{C \neq D} \mathbb{E}[\mathbf{I}_C \mathbf{I}_D] - \sum_{C \neq D} \mathbb{E}[\mathbf{I}_C]\mathbb{E}[\mathbf{I}_D] \\ &= \sum_C \mathrm{Var}(\mathbf{I}_C) + \sum_{C \neq D} \mathrm{Cov}(\mathbf{I}_C, \mathbf{I}_D) \ . \end{aligned} \tag{2.28}$$

Since the $\mathbf{I}_C$ are indicator random variables, so only take values equal to zero or one, we have $\sum_C \mathrm{Var}(\mathbf{I}_C) = \sum_C \left(\mathbb{E}[\mathbf{I}_C] - \mathbb{E}[\mathbf{I}_C]^2\right) = \Theta(n^4 p^6)$. To evaluate the covariance cases, note that each K_4 can share either zero, one, two or three vertices with another K_4. In the case when they share zero or one vertices, their presence in $G(n,p)$ is independent of one another, so do not contribute to the covariance term. If C and D share exactly two vertices, the total contribution of this covariance term is $\Theta\left(n^6 p^{11}\right)$. In the case where C and D share three vertices, the contribution of this covariance term is $\Theta\left(n^5 p^9\right)$. Putting all these into equation (2.28), we hence obtain

$$\mathrm{Var}(S_n) = \Theta\left(n^4 p^6\right) + \Theta\left(n^6 p^{11}\right) + \Theta\left(n^5 p^9\right) = \Theta\left(n^6\right)$$

in the case when p is constant, as required. ■

2.5.1 Central Limit Theorem for Number of K_4

By using a moment method one can prove a central limit theorem for the number of K_4 subgraphs of $G(n,p)$ for the fixed edge probability case, as in [20, Theorem 3.3.6.].

Theorem 20. ***(Central limit theorem for the number of*** $\mathbf{K_4}$ ***in*** $\mathbf{G(n,p)}$***)*** *For fixed constants a, b the following holds*

$$\left|\mathbb{P}\left(a \leqslant \frac{\mathbf{S}_n - \mu_n}{\sigma_n} \leqslant b\right) - \int_a^b \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt\right| = o(1) \ . \tag{2.29}$$

In other words, $\frac{\mathbf{S}_n - \mu_n}{\sigma_n}$ converges in distribution to the standard Gaussian distribution.

2.5.2 Chernoff Bounds

Many of the following arguments will make use of Chernoff bounds. Here we give a quick overview of these Chernoff bounds.

Theorem 21. *[19]* ***(Two-sided Binomial Chernoff bound)*** *Let $\mathbf{X}$ be a Binomial random variable with parameters n and p. For any $0 \leqslant t \leqslant np$ we have,*

$$\mathbb{P}(|\mathbf{X} - np| \geq t) \leqslant 2\exp\left\{-\frac{t^2}{3np}\right\} \ . \tag{2.30}$$

Theorem 22. *[15, Theorem 2]* ***(General Chernoff bound)*** *Suppose that $\{\mathbf{X}_i\}_{i=1}^n$ is a collection of independent random variables such that $X_i \in [a_i, b_i]$. Further, let $\mathbf{X}$ denote their sum, so $\mathbf{X} = \sum_{i=1}^n \mathbf{X}_i$, and let μ denote the mean of $\mathbf{X}$, so $\mu = \mathbb{E}[\mathbf{X}] = \sum_{i=1}^n \mathbb{E}[\mathbf{X}_i]$. For $t \geqslant 0$*

$$\mathbb{P}(|\mathbf{X} - \mu| \geq t) \leqslant 2\exp\left\{-\frac{2t^2}{n\sum_{i=1}^n (b_i - a_i)^2}\right\} \ . \tag{2.31}$$

2.6 Outline of Main Conditioning Argument

To prove the local limit theorem, we need to show that for the ranges R_2-R_7 the absolute value of the characteristic function of the number of copies of K_4 is not too big. We show this by using different conditioning arguments. We will outline the general structure of these now. Note that region R_1 is being taken care of by the Central limit theorem Theorem 20, so bounding the characteristic function in ranges R_2-R_7 suffices to prove our result. The remainder of this subsection will not be fully rigorous in places and is meant mainly for the purpose of giving a intuitive idea of the argument used in the following subsections.

We will aim to bound the absolute value of the characteristic function by using the following trick: Suppose $\mathbf{S}_n$ was the sum of k independent identically distributed random variables $\{X_i\}_{i=1}^k$, so

$$\mathbf{S}_n = \sum_{i=1}^k X_i \ .$$

Then we would have

$$\begin{aligned}\psi_n(\theta\sigma_n) &= \mathbb{E}\left[e^{i\theta\sum_{i=1}^{k} X_i}\right] \\ &= \prod_{i=1}^{k} \mathbb{E}\left[e^{i\theta X_i}\right] \text{ since the } X_i \text{ are independent} \\ &= \mathbb{E}\left[e^{i\theta X_1}\right]^k \text{ since the } X_i \text{ are identically distributed.}\end{aligned}$$

So to show that the absolute value of the function ψ_n is small, we just need to show that $\left|\mathbb{E}\left[e^{i\theta X_i}\right]\right|$ is sufficiently bounded away from 1. Unfortunately, when counting the number of K_4, we clearly do not have independent random variables.

However, we will condition on some subset of the edges, F say, of $G(n,p)$ and show that with high probability each of the k remaining edges is adjacent to around x copies of K_4^- that are entirely contained in the set of edges we have conditioned on.

As an example, consider the K_3 case. If we let E be a matching of size $\lfloor \frac{n}{2} \rfloor$ and $F = E(G) \setminus E$. Then each edge $e \in E$ is adjacent on average to np^2 copies of K_3^- in $G(n,p)$. So by standard concentration bounds and independence, each edge $e \in E$ is with high probability adjacent to roughly np^2 copies of K_3^-. So conditioning on F and conditional on this event happening, $\mathbf{S}_n$ is the sum of $|E| = \frac{n}{2}$ random variables each multiplied by a constant of size roughly np^2. Thus

$$\begin{aligned}\mathbb{E}[e^{i\theta\mathbf{S}_n}] &\approx \mathbb{E}[e^{i\theta\sum_{e\in E} c_e\mathbf{X}_e}] \\ &= \prod_{e\in E} \mathbb{E}[e^{i\theta c_e\mathbf{X}_e}] \\ &\approx \mathbb{E}[e^{i\theta c_e\mathbf{X}_e}]^{n/2} \\ &\leqslant \left(1 - 8p(1-p)\left|\left|\frac{\theta np^2}{2\pi}\right|\right|^2\right)^{n/2} \\ &\leqslant \left(1 - 8p(1-p)\left(\frac{\theta np^2}{2\pi}\right)^2\right)^{n/2} \text{ as long as } \theta np^2 < \pi \\ &\leqslant e^{-\Theta(\theta^2 n^3)} ,\end{aligned}$$

where we have written $\mathbf{X}_e$ for the indicator random variable for the presence of the edge e and c_e for the number of K_3^- in F the edge e is adjacent to. However, note that it is impossible to achieve that every edge e is only adjacent to K_4^- entirely contained in F unless F contains all but one edge. We will, in the K_4 case thus always have to also consider some cross terms. To deal with these, we show that the influence of the cross-terms, so the K_4 containing more than one of the edges inside the k remaining edges, the set E say, is sufficiently small so that the absolute value of $\psi_n(\theta\sigma_n)$ can still be bound above by something small enough. Denote the cross terms by η. We would

then be left with k independent, identically distributed Bernoulli random variables with success parameter p, each of scale x. So conditional on some events that occur with high probability, we can bound ψ_n as follows:

$$\begin{aligned}
\psi_n(\theta\sigma_n) &\leqslant \mathbb{E}\left[e^{i\theta x\sum_{i=1}^{k} X_i}\right] + \eta \\
&= \prod_{i=1}^{k} \mathbb{E}\left[e^{i\theta x X_i}\right] + \eta \\
&= \mathbb{E}\left[e^{i\theta x X_i}\right]^k + \eta \\
&\leqslant \left(1 - 8p(1-p)\left|\left|\frac{\theta x}{2\pi}\right|\right|^2\right)^k + \eta \quad \text{by Lemma 14} \\
&\leqslant \left(1 - 8p(1-p)\left(\frac{\theta x}{2\pi}\right)^2\right)^k + \eta \quad \text{as long as } \theta x < \pi \\
&\leqslant e^{-\Theta(\theta^2 x^2 k)} + \eta \ .
\end{aligned}$$

So as long as $\theta^2 x^2 k = \omega(\log(n))$, and η is also small enough, we can bound the absolute value of the characteristic function above by $(\theta\sigma)^{-(1+\delta)}$ to get the desired local convergence result. To cover the entire range of integration for θ, we need different combinatorial constructions and conditioning arguments to give us different scales x. Recall that for a given construction, we have with high probability that the presence of one of at least a constant fraction of the edges inside E makes a difference of size around x on the total number of copies of K_4 inside $G(n,p)$. So if the scale is of order x, we can use this conditioning argument to bound the characteristic function sufficiently for θ around $\omega\left(\frac{\sqrt{\log(n)}}{x\sqrt{k}}\right) = \theta < \frac{\pi}{x}$ provided that the contribution of the cross terms is small enough.

2.7 Proof of Main Theorem

2.7.1 Region $R_1 = (0, A/\sigma_n)$

We now prove Lemma 7. Note that this follows from the central limit theorem for the number of K_4 in $G(n,p)$ [20]. Let G be an arbitrary graph with at least one edge and p be a constant in $(0,1)$. The central limit theorem, Theorem 2 in [20], then states that the number of copies of G in $G(n,p)$, denoted by $\mathbf{S}_n$, satisfies

$$\frac{\mathbf{S}_n - \mathbb{E}[\mathbf{S}_n]}{\sqrt{\mathrm{Var}(\mathbf{S}_n)}} \xrightarrow{\mathscr{D}} \mathcal{N}(0,1) \ , \tag{2.32}$$

where $\mathscr{D}$ denotes convergence in distribution. Note that in [5, Theorem 3.3.6.] Durrett gives the following theorem relating convergence in distribution to point-wise convergence of characteristic functions.

Theorem 23. *[5, Theorem 3.3.6.]* ***(Continuity Theorem)***
Let $(\mathbf{X}_n)_{n=1}^{\infty}$ *be sequence of random variables with characteristic functions* $\phi_n(x)$ *and* $\mathbf{X}$ *another random variable with characteristic function* $\phi(x)$*. If* $\mathbf{X}_n$ *convergence to* $\mathbf{X}$ *in distribution, then* $\phi_n(x) \to \phi(x)$ *for all* x*.*

Combining Theorem 23 with equation (2.32), we get that $\psi_n(t) \to e^{-t^2/2}$ point-wise for all t. By the dominated convergence theorem, we thus have that for any fixed constant $A \in \mathbb{R}$

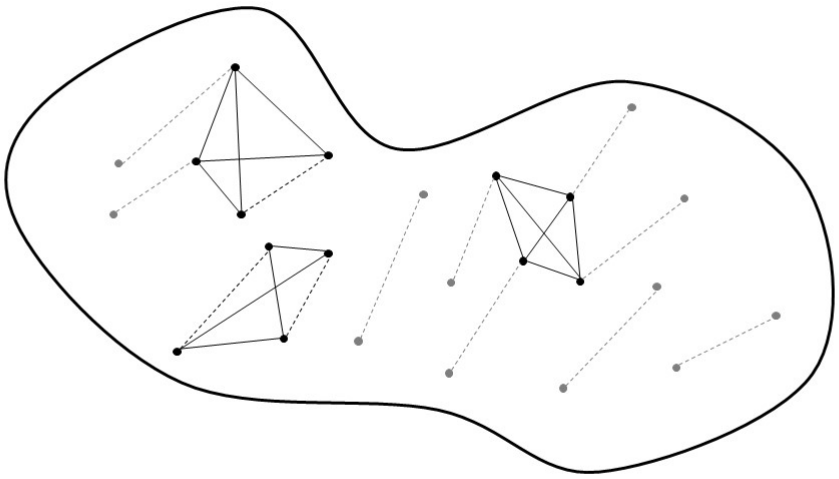
$$\int_0^A \left| \psi_n(t) - e^{-t^2/2} \right| dt \to 0 \ , \tag{2.33}$$

which proves Lemma 7.

2.7.2 Region $R_3 = \left(n^{-2.45}, n^{-2.1}\right)$

We now prove Lemma 9, using a conditioning argument similar to the one outlined in Section 2.6.

Let E be a set of k edges forming a matching inside K_n, so a set of possible edges in $G(n,p)$. We first reveal all edges in $F = \binom{[n]}{2} \setminus E$ and later condition on these.



Then $\mathbf{S}_n$ can be written as

$$\mathbf{S}_n = \mathbf{C}(F) + \mathbf{Y}(E,F) + \mathbf{Z}(E,F) \ ,$$

where $\mathbf{C}(F)$, $\mathbf{Y}(E,F)$ and $\mathbf{Z}(E,F)$ are the number of K_4 subgraphs in $G(n,p)$ with zero, one or two edges in the set of possible edges E respectively. Letting $\alpha = \mathbb{E}_{E,F}[Z(E,F)]$

and applying Lemma 16 we have

$$\begin{aligned}\left|\mathbb{E}\left[e^{i\theta \mathbf{S}_n}\right]\right| &\leqslant \left|\mathbb{E}\left[e^{i\theta\,(\mathbf{C}(F)+\mathbf{Y}(E,F))}\right]\right| + \mathbb{E}\left[\left|e^{i\theta\,(\mathbf{Z}(E,F)-\alpha)}-1\right|\right] \\ &= \left|\mathbb{E}_F\left[e^{i\theta\,\mathbf{C}(F)}\,\mathbb{E}_E\left[e^{i\theta\,\mathbf{Y}(E,F)}\right]\right]\right| + \mathbb{E}\left[\left|e^{i\theta\,(\mathbf{Z}(E,F)-\alpha)}-1\right|\right] \\ &\leqslant \mathbb{E}_F\left[\left|e^{i\theta\,\mathbf{C}(F)}\,\mathbb{E}_E\left[e^{i\theta\,\mathbf{Y}(E,F)}\right]\right|\right] + \mathbb{E}\left[\left|e^{i\theta\,(\mathbf{Z}(E,F)-\alpha)}-1\right|\right] \\ &\leqslant \mathbb{E}_F\left[\left|\mathbb{E}_E\left[e^{i\theta\,\mathbf{Y}(E,F)}\right]\right|\right] + \mathbb{E}\left[\left|e^{i\theta\,(\mathbf{Z}(E,F)-\alpha)}-1\right|\right]\ . \end{aligned} \tag{2.34}$$

So to bound $\left|\mathbb{E}\left[e^{i\theta \mathbf{S}_n}\right]\right|$, it is enough to show that the two expectations on the right hand side are both small in absolute value. We bound these two expectations separately in the following two subsections.

2.7.2.1 Bounding $\mathbb{E}_F\left[\left|\mathbb{E}_E\left[e^{i\theta\,\mathbf{Y}(E,F)}\right]\right|\right]$

To bound $\mathbb{E}_F\left[\left|\mathbb{E}_E\left[e^{i\theta\,\mathbf{Y}(E,F)}\right]\right|\right]$, it suffices to show that for most choices of F, the expectation $\mathbb{E}_E\left[e^{i\theta\,\mathbf{Y}(E,F)}\right]$ is small in absolute value. For the remainder of this section, we will use the more formal notation of $\mathbf{X}(X_E)$ to denote that the random variable $\mathbf{X}$ depends on the presence or absence of edges in E. We denote by x_E an observation of X_E.

For $e=\{u,v\}\in E$ and a vector $x_F\in\{0,1\}^F$, we let $\mathbf{Y}_e(x_F)$ be the number of K_4^- based at $e=\{u,v\}$ and consisting entirely of edges $f\in F$ for which $(x_F)_f=1$. Let $\mathbf{X}_e$ be the indicator random variable of the presence of the edge e in G. Then

$$\mathbf{Y}(X_E,x_F)=\sum_{e\in E}\mathbf{Y}_e(x_F)\cdot\mathbf{X}_e\ . \tag{2.35}$$

We define

$$\mathbf{L}=\left\{x_F\in\{0,1\}^F\ \middle|\ \text{for some } e\in E \text{ we have } \mathbf{Y}_e(x_F)<\frac{n^2p^5}{10}\right\}\ . \tag{2.36}$$

Let Λ_e be the "bad" event that the number of K_4^- based at e and entirely contained in F is less than $\frac{n^2p^5}{10}$. We further define Λ to be the ("bad") event that $\mathbf{X}_F\in L$, so $\Lambda=\bigcup_{e\in E}\Lambda_e$.

To bound the probability of Λ occurring, we show that for each edge e another event Γ_e^{c} occurs with high probability and conditional on Γ_e^{c}, the event Λ_e^{c} also occurs with high probability. This then allows us to bound the probability of $\Lambda=\bigcup_{e\in E}\Lambda_e$. For a given edge $e=\{u,v\}$, let m_e count the number of common neighbours of u and v. Note that m_e is distributed as a binomial random variable with parameters $(n-2)$ and p^2. We let Γ_e be the "bad" event that $m_e<\frac{(n-2)p^2}{2}$.

Lemma 24. *For any edge $e\in E$ we have that $\mathbb{P}(\Gamma_e)\leqslant e^{-\Theta(n)}$.*

Proof. Since the number of common neighbours of e inside F is binomially distributed with parameters $(n-2)$ and p^2, we have by applying the Chernoff bound from Theorem 21 that

$$\begin{aligned}\mathbb{P}(\Gamma_e) &= \mathbb{P}\left(\mathrm{Bin}((n-2),p^2) < \frac{(n-2)p^2}{2}\right)\\ &\leqslant \exp\left\{-\frac{(n-2)p^2}{8}\right\}\\ &= e^{-\Theta(n)}\ ,\end{aligned}$$

as required. ■

Lemma 25. *We have that* $\mathbb{P}_F(\Lambda) \leqslant ke^{-\Theta(n)}$.

Proof. We now want to show that, conditional on the event Γ_e not occurring, the number of K_4^- based at $e = \{u,v\} \in E$ and entirely contained in F is with high probability at least $\frac{n^2p^5}{10}$. Since we have shown that the probability that Γ_e occurs is small, this is enough to give us the required result. For an edge $e \in E$, denote by x_e the number of pairs $\{a,b\}$ of common neighbours of e such that $\{a,b\} \in E$, so x_e counts the number of possible K_4^- based at e but not entirely contained in F. Note that given m_e and x_e, the number of K_4^- based at u and v and entirely contained inside F is binomially distributed with parameters n_e and p, where $n_e = \binom{m_e}{2} - x_e$. Since E contains k edges, we know that $x_e \leqslant k$. Note further that conditioning on the event Γ_e^{c} means conditioning on m_e being at least $\frac{(n-2)p^2}{2}$. The probability of Λ_e is then equal to the probability of the number of K_4^- based at e being less than $\frac{n^2p^5}{10}$. When Γ_e^{c} holds n_e satisfies [1]

$$n_e \geq \binom{\frac{(n-2)p^2}{2}}{2} - k \geq \frac{n^2p^4}{9}$$

since $k \leqslant n$. By the binomial Chernoff bound from Theorem 21

$$\begin{aligned}\mathbb{P}(\Lambda_e \mid \Gamma_e^{\mathrm{c}}) &= \mathbb{P}\left(\mathrm{Bin}(n_e,p) < \frac{n^2p^5}{10} \mid \Gamma_e^{\mathrm{c}}\right)\ ,\\ &\leqslant e^{-\Theta\left(n^2\right)}\ .\end{aligned}$$

Since we also have $\mathbb{P}(\Gamma_e) \leqslant e^{-\Theta(n)}$, it follows that

$$\mathbb{P}(\Lambda_e) \leqslant e^{-\Theta(n)}\ , \tag{2.37}$$

and now applying a union bound over all $e \in E$ we have

$$\mathbb{P}(\Lambda) \leqslant ke^{-\Theta(n)}\ , \tag{2.38}$$

as required. ■

[1] Here and where necessary in what follows, we will assume that n is sufficiently large

Next we show that the expectation $\mathbb{E}_E\left[e^{i\theta\, Y(X_E,x_F)}\right]$ is small in absolute value for any choice of $x_F \notin \mathbf{L}$.

Lemma 26. *For every $x_F \in \{0,1\}^F \setminus \mathbf{L}$ we have*

$$\left|\mathbb{E}_E\left[e^{i\theta\, Y(X_E,x_F)}\right]\right| \leqslant e^{-\Theta\left(\theta^2 n^4 k\right)} \ . \tag{2.39}$$

Proof. Recall that $\mathbf{Y}(X_E, x_F) = \sum_{e\in E} \mathbf{Y}_e(x_F)\cdot \mathbf{X}_e$, so

$$\begin{aligned}
\left|\mathbb{E}_{X_E}\left[e^{i\theta\,\mathbf{Y}(X_E,x_F)}\right]\right| &= \left|\mathbb{E}_{X_E}\left[e^{i\theta\,\sum_{e\in E}\mathbf{Y}_e(x_F)\cdot\mathbf{X}_e}\right]\right| \\
&= \left|\prod_{e\in E}\mathbb{E}_{X_E}\left[e^{i\theta\,\mathbf{Y}_e(x_F)\cdot\mathbf{X}_e}\right]\right| \text{ by independence of the } (\mathbf{X}_e)_{e\in E} \\
&\leqslant \prod_{e\in E}\left(1-8p(1-p)\left|\left|\frac{\theta\,\mathbf{Y}_e(x_F)}{2\pi}\right|\right|^2\right) \text{ by Lemma 14} \\
&\leqslant \left(1-8p(1-p)\left(\frac{\theta\,\mathbf{Y}_e(x_F)}{2\pi}\right)^2\right)^k \text{ since } \mathbf{Y}_e \leqslant \frac{n^2}{2} \text{ and } \theta < 2\pi n^{-2} \\
&\leqslant \left(1-8p(1-p)\left(\frac{\theta n^2 p^5}{20\pi}\right)^2\right)^k \text{ since } x_F \notin \mathbf{L} \\
&\leqslant e^{-8p(1-p)\left(\frac{\theta n^2 p^5}{20\pi}\right)^2\cdot k} \\
&\leqslant e^{-\Theta\left(\theta^2 n^4 k\right)} \ ,
\end{aligned}$$

as required. ■

Combining Lemma 26 and Lemma 25 we get the following bound on the expectation

$$\begin{aligned}
\mathbb{E}_F\left[\left|\mathbb{E}_E\left[e^{i\theta\,\mathbf{Y}(X_E,X_F)}\right]\right|\right] &= \mathbb{P}(X_F\in\mathbf{L}) + \max_{x_F\in\{0,1\}^F\setminus L}\left|\mathbb{E}_E\left[e^{i\theta\cdot\mathbf{Y}(X_E,X_F)}\right]\right| \\
&\leqslant ke^{-\Theta(n)} + e^{-\Theta\left(\theta^2 n^4 k\right)} \ .
\end{aligned} \tag{2.40}$$

2.7.2.2 Bounding $\mathbb{E}\left[\left|e^{i\theta\cdot(\mathbf{Z}(E,F)-\alpha)}-1\right|\right]$

To bound this, we want to show that $\mathbf{Z}(E,F)$ is close to its expectation $\alpha = \mathbb{E}\left[\mathbf{Z}(E,F)\right]$ with high probability, which reduces to proving a concentration bound. We do this by bounding the variance of $\mathbf{Z}(E,F)$. Denote by Γ the set of all K_4 in K_n that have exactly two edges in E. Furthermore, for $r\in\Gamma$, let $T_r(E,F)$ be the indicator random variable for this copy of K_4 appearing in G. For any two $r,s\in\Gamma$, we write $r\sim s$ if they share an edge. Note that any $s\in\Gamma$ with $s\sim r$ can either share one of the edges in F or one of the edges in E with r. If they share one of the edges in E, this edge fixes two of the vertices in s and the other two vertices are connected to each other by an edge in E. Thus for a fixed K_4, there are at most $2(k-2)$ other K_4 sharing exactly one of the two edges in

E with it. If r and s share an edge in F, f say, then the end-vertices of f are adjacent to exactly one edge in E each. Since E is a matching, this fixes the other two vertices in the K_4, so $s = r$. We thus have that for any $r \in \Gamma$, r shares an edge with at most $2k$ copies s of K_4 such that $s \in \Gamma$. Thus

$$\begin{aligned}
\mathrm{Var}(\mathbf{Z}(E,F)) &= \sum_{r\in\Gamma}\sum_{s\in\Gamma} \mathrm{Cov}_{E,F}(T_r(E,F), T_s(E,F)) \\
&= \sum_{r\in\Gamma}\sum_{s\sim r} \mathrm{Cov}_{E,F}(T_r(E,F), T_s(E,F)) \text{ by independence} \\
&\leqslant \binom{k}{2} \cdot 2k \\
&\leqslant k^3 \ .
\end{aligned}$$

By Lemma 15, we thus have

$$\mathbb{E}\left[\left|e^{i\theta\cdot[\mathbf{Z}(E,F)-\alpha]} - 1\right|\right] \leqslant \theta\sqrt{\mathrm{Var}\left(\mathbf{Z}(E,F)-\alpha\right)} \leqslant \theta k^{\frac{3}{2}} \ . \tag{2.41}$$

2.7.2.3 Combining Bounds

We now prove Lemma 9. Combining equations (2.34), (2.40) and (2.41) we have

$$\begin{aligned}
\left|\mathbb{E}\left[e^{i\theta\mathbf{S}_n}\right]\right| &\leqslant \left|\mathbb{E}\left[e^{i\theta\left(\mathbf{C}(F)+\mathbf{Y}(E,F)\right)}\right]\right| + \mathbb{E}\left[\left|e^{i\theta\left(\mathbf{Z}(E,F)-\alpha\right)} - 1\right|\right] \\
&\leqslant ke^{-\Theta(n)} + e^{-\Theta\left(\theta^2 n^4 k\right)} + \theta k^{\frac{3}{2}} \ .
\end{aligned} \tag{2.42}$$

We now choose $k = \lfloor \theta^{-2} n^{-3.92} \rfloor$. We claim that we can bound $\left|\mathbb{E}\left[e^{i\theta\mathbf{S}_n}\right]\right|$ above by $(\theta\sigma_n)^{-(1+\delta)}$, where $\delta = 0.45$. To complete the proof, we need to check that the the following two conditions hold for all θ such that $n^{-2.45} \leqslant \theta \leqslant n^{-2.1}$:

1. equation (2.42) is bounded above by $(\theta\sigma_n)^{-1.45}$
2. k is sensible, in other words lies in the interval $\{1, \ldots, n/2\}$

In order for condition 1 to hold, note that it suffices if each term of the sum is bound above by $1/3(\theta\sigma_n)^{-1.45}$, which holds as long as

$$\omega(\theta^{-2}n^{-4}\log(n)) \leqslant k \leqslant \theta^{-1.64}n^{2.9} \ .$$

Our choice of $k = \lfloor \theta^{-2} n^{-3.92} \rfloor$ satisfies this, so condition 1 holds.

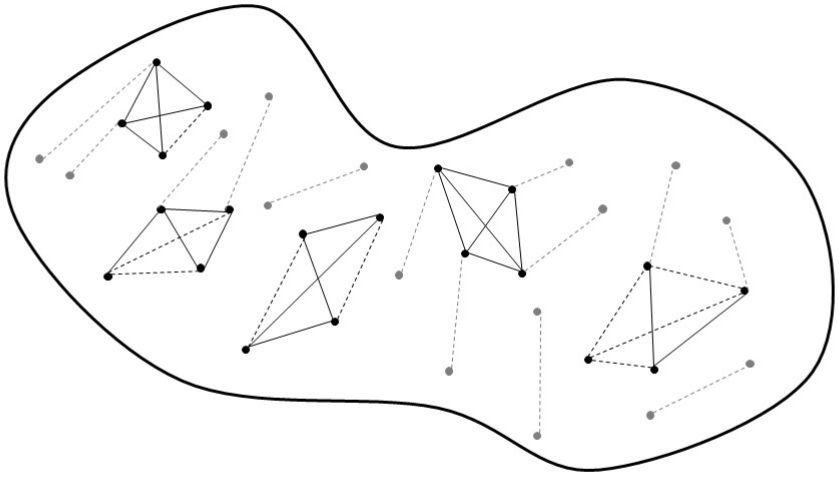
To establish condition 2 note that for $n^{-2.45} \leqslant \theta \leqslant n^{-2.1}$ we have that

$$\begin{aligned}
\lfloor \theta^{-2} n^{-3.92} \rfloor &\leqslant \lfloor n^{4.9} n^{-3.92} \rfloor = \lfloor n^{0.98} \rfloor < n/2 \\
\lfloor \theta^{-2} n^{-3.92} \rfloor &\geq \lfloor n^{4.2} n^{-3.92} \rfloor = \lfloor n^{0.28} \rfloor > 1
\end{aligned}$$

for n sufficiently large. Thus condition 2. is also satisfied for our choice of k as required. This proves Lemma 9.

2.7.3 Region $R_2 = \left(0, n^{-2.45}\right)$

We now prove Lemma 8. Note that trivially $|\psi_n(\theta)| \leqslant 1$ and so the lemma already holds for $\theta \leqslant D_2^{3/2}\sigma_n^{-1}$, where D_2 is a constant to be defined later. We may thus assume that θ is bigger than $D_2^{3/2}/\sigma_n$ and n is sufficiently large. We prove the bound in this region R_2 again by first revealing a subset of the edges in $G(n,p)$ and conditioning on these similarly to before.



Let E be the set of all edges in $l = \left\lfloor \frac{n}{2k} \right\rfloor$ copies of $K_{k,k}$ inside K_n, where $1 \leqslant l = o(n)$. We first reveal all edges in $F = \binom{[n]}{2} \setminus E$ and condition on these edges. Then

$$\mathbf{S}_n = \mathbf{C}(F) + \mathbf{W}(E,F) + \mathbf{Y}(E,F) + \mathbf{Z}(E,F) \ ,$$

where $\mathbf{C}(F)$, $\mathbf{W}(E,F)$, $\mathbf{Y}(E,F)$ and $\mathbf{Z}(E,F)$ are the number of K_4 subgraphs in $G(n,p)$ with zero, one, two or at least three edges in E, respectively. Letting $\alpha = \mathbb{E}_{E,F}\left[\mathbf{Y}(E,F)\right]$ and $\beta = \mathbb{E}_{E,F}\left[\mathbf{Z}(E,F)\right]$ and applying Lemma 16 we have

$$\begin{aligned} \left|\mathbb{E}\left[e^{i\theta\mathbf{S}_n}\right]\right| &= \left|\mathbb{E}\left[e^{i\theta(\mathbf{C}(F)+\mathbf{W}(E,F)+\mathbf{Y}(E,F)+\mathbf{Z}(E,F))}\right]\right| \\ &\leqslant \left|\mathbb{E}\left[e^{i\theta(\mathbf{C}(F)+\mathbf{W}(E,F))}\right]\right| + \mathbb{E}\left[\left|e^{i\theta(\mathbf{Y}(E,F)-\alpha)} - 1\right|\right] + \mathbb{E}\left[\left|e^{i\theta(\mathbf{Z}(E,F)-\beta)} - 1\right|\right] \ . \end{aligned} \tag{2.43}$$

So to bound $\left|\mathbb{E}\left[e^{i\theta\mathbf{S}_n}\right]\right|$, it suffices to show that each of the three expectations is sufficiently small. We bound the three expectations separately in the following subsections.

2.7.3.1 Bounding $\left|\mathbb{E}\left[e^{i\theta(\mathbf{C}(F)+\mathbf{W}(E,F))}\right]\right|$

Note that

$$\left|\mathbb{E}_{E,F}\left[e^{i\theta(\mathbf{C}(F)+\mathbf{W}(E,F))}\right]\right| \leqslant \mathbb{E}_F\left[\left|\mathbb{E}_E\left[e^{i\theta\,\mathbf{W}(E,F)}\right]\right|\right] \ .$$

So to bound $\left|\mathbb{E}\left[e^{i\theta(\mathbf{C}(F)+\mathbf{W}(E,F))}\right]\right|$, it suffices to show that $\mathbb{E}_E\left[e^{i\theta\,\mathbf{W}(E,F)}\right]$ is on average small in absolute value. For $e=\{u,v\}\in E$ and a vector $x_F\in\{0,1\}^F$, let $\mathbf{W}_e(x_F)$ denote the number of K_4^- based at $\{u,v\}$ consisting entirely of edges $f\in F$ for which $(x_F)_f=1$, and let $\mathbf{X}_e$ be the indicator random variable for the presence of the edge e. Then

$$\mathbf{W}(X_E,x_F)=\sum_{e\in E}\mathbf{W}_e(x_F)\cdot\mathbf{X}_e \ . \tag{2.44}$$

Let

$$\mathbf{L}=\left\{x_F\in\{0,1\}^F \mid \text{ for some } e\in E \text{ we have } W_e(x_F)<\frac{n^2p^5}{10}\right\} \ . \tag{2.45}$$

For the range of θ we are considering in this section, we want the presence of an edge in E to change the number of K_4 in $G(n,p)$ by $\Theta(n^2)$. We let Λ_e be the "bad" event that the number of K_4^- based at e and entirely contained in F is less than $\frac{n^2p^5}{10}$. Further, let Λ be the "bad" event that $X_F\in\mathbf{L}$, so $\Lambda=\bigcup_{e\in E}\Lambda_e$ and aim to show that the probability of Λ occurring is small. We again show that the probability of Λ occurring is small by showing that for each of the edges e, another event Γ_e^{c} occurs with high probability. Conditional on this event, the event Λ_e^{c} also occurs with high probability. This then allows us to bound the probability of $\Lambda=\bigcup_{e\in E}\Lambda_e$. As before, for an edge $e=\{u,v\}\in E$, let Γ_e be the the event that the number m_e of vertices that both u and v are connected to through edges in F is less than $\frac{(n-2k)p^2}{2}$.

Lemma 27. *We have that* $\mathbb{P}(\Gamma_e)\leqslant e^{-\Theta(n)}$.

Proof. The proof is similar to the proof of Lemma 24, however in this case in U and V we have $n-2k$ potential common neighbours in F, which is still of order n as k is small enough. ■

Lemma 28. *We have that* $\mathbb{P}(\Lambda)\leqslant nke^{-\Theta(n)}$.

Proof. The proof is the same as for Lemma 25, with the slight difference that this time $x_e\leqslant 2nk$. Note however that since $k=o(n)$, we still have when Γ_e holds

$$n_e\geq\binom{\frac{(n-2k)p^2}{2}}{2}-2nk\geqslant\frac{n^2p^4}{9} \tag{2.46}$$

and so the weaker bound on x_e does not affect the bound on the probability of Λ_e. The only other difference to the proof of Lemma 25 is that in the final step we are taking the union over nk rather than just k events, so $\mathbb{P}(\Lambda)\leqslant nke^{-\Theta(n)}$ as required. ■

Next we show that if we condition on Λ not occurring, the desired expectation is small.

Lemma 29. *For a fixed set of edges $x_F \notin \mathbf{L}$, we have*

$$\left|\mathbb{E}_E\left[e^{i\theta\,\mathbf{W}(X_E,x_F)}\right]\right| \leqslant e^{-\Theta\left(\theta^2 n^5 k\right)}\ . \tag{2.47}$$

Proof. Recall that $\mathbf{W}(X_E, x_F) = \sum_{e\in E}\mathbf{W}_e(x_F)\cdot\mathbf{X}_e$, so

$$\begin{aligned}
\left|\mathbb{E}_E\left[e^{i\theta\,\mathbf{W}(X_E,x_F)}\right]\right| &= \left|\mathbb{E}_E\left[e^{i\theta\,\sum_{e\in E}(\mathbf{W}_e(x_F)\cdot\mathbf{X}_e)}\right]\right| \\
&= \left|\prod_{e\in E}\mathbb{E}_E\left[e^{i\theta\,\mathbf{W}_e(x_F)\cdot\mathbf{X}_e}\right]\right| \text{ by independence of } (\mathbf{X}_e)_{e\in E} \\
&\leqslant \prod_{e\in E}\left(1-8p(1-p)\left|\left|\frac{\theta\,\mathbf{W}_e(x_F)}{2\pi}\right|\right|^2\right) \text{ by Lemma 14} \\
&\leqslant \left(1-8p(1-p)\left(\frac{\theta\,\mathbf{W}_e(x_F)}{2\pi}\right)^2\right)^{nk} \text{ since } \mathbf{W}_e \leqslant \frac{n^2}{2},\ \theta < 2\pi n^{-2} \\
&\leqslant \left(1-8p(1-p)\left(\frac{\theta}{2\pi}\cdot\frac{n^2p^5}{10}\right)^2\right)^{nk} \text{ since } x_F \notin \mathbf{L} \\
&\leqslant e^{-8p(1-p)\left(\frac{\theta n^2p^5}{20\pi}\right)^2 nk} \\
&\leqslant e^{-\Theta\left(\theta^2 n^5 k\right)}\ ,
\end{aligned}$$

as required. ■

Combining Lemmas 28 and 29 we obtain

$$\begin{aligned}
\left|\mathbb{E}_{E,F}\left[e^{i\theta\,\mathbf{W}(X_E,X_F)}\right]\right| &\leqslant \mathbb{P}_F(X_F\in\mathbf{L}) + \max_{x_F\in\{0,1\}^F\setminus\mathbf{L}}\left|\mathbb{E}_E\left[e^{i\theta\cdot\mathbf{W}(X_E,x_F)}\right]\right| \\
&\leqslant nke^{-\Theta(n)} + e^{-\Theta\left(\theta^2 n^5 k\right)}\ .
\end{aligned} \tag{2.48}$$

2.7.3.2 Bounding $\mathbb{E}\left[\left|e^{i\theta(\mathbf{Y}(E,F)-\alpha)}-1\right|\right]$

To bound the absolute value of this expectation, we want to show that $\mathbf{Y}(E,F)$ is close to its expectation $\alpha = \mathbb{E}[\mathbf{Y}(E,F)]$ with high probability. As in Section 2.7.2.2, we do so by bounding the variance of $\mathbf{Y}(E,F)$. Denote by Γ_2 the set of all K_4 in K_n that have exactly two edges in E. As before, for $r\in\Gamma$, let $T_r(E,F)$ be the indicator random variable for this copy of K_4 appearing in G. For any two $r,s\in\Gamma$, we again write $r\sim s$ if they share an edge.

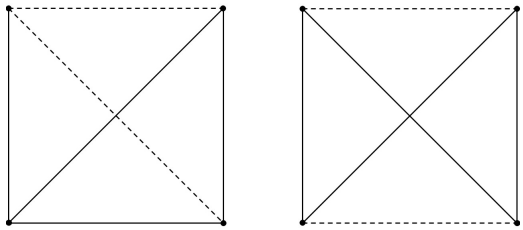


Figure 2.1: The only possible types of K_4 in Γ_2

In order to determine $|\Gamma_2|$, as well as the number of pairs r, s in Γ_2 such that $r \sim s$, we now count the number of r in Γ_2 that an edge in E or F can be contained in. Suppose e is an edge in E and e is contained in r in Γ_2. Then e can either be adjacent to another edge in E in $r \in \Gamma_2$ or not, as seen in Figure 2.1. If it is, there are at most $2k$ choices for the second edge in E and then there are at most n choices for the final vertex in r. If $e \in E$ is not adjacent to another edge in E in r, then there are $\leqslant nk$ choices for the second edge in E in r and this determines r completely. Thus an edge $e \in E$ is contained in $\leqslant 3nk$ copies $r \in \Gamma_2$.

Now suppose $f \in F$ is contained in $r \in \Gamma_2$. Then f is adjacent to at least one edge $e \in E$ in r. There are at most $2k$ choices for this. Then, either e or f is adjacent to another edge in E in $r \in \Gamma_2$ whose other endvertex is not adjacent to e or f; or f is adjacent to an edge in F in $r \in \Gamma_2$ whose other endvertex is not adjacent to e or f. There are at most $3k + n$ choices for this and this determines $r \in \Gamma_2$ completely. Thus an edge $f \in F$ is contained in total in at most $6k^2 + 2nk \leq 3nk$ copies $r \in \Gamma_2$.

For a given $r \in \Gamma_2$ there are as a result at most $2 \cdot 3nk + 4 \cdot 3nk \leqslant 18nk$ copies $s \in \Gamma_2$ such that $r \sim s$. Since there are at most nk edges in E and each of these is contained in at most $3nk$ copies $r \in \Gamma_2$, we have that $|\Gamma_2| \leqslant 6n^2k^2$. Thus

$$\begin{aligned}
\mathrm{Var}(\mathbf{Y}(E,F)) &= \sum_{r\in\Gamma_2}\sum_{s\in\Gamma_2} \mathrm{Cov}_{E,F}(T_r(E,F), T_s(E,F)) \\
&= \sum_{r\in\Gamma_2}\sum_{s\sim r} \mathrm{Cov}_{E,F}(T_r(E,F), T_s(E,F)) \text{ by independence} \\
&\leqslant 6n^2k^2 \cdot 18nk \\
&\leqslant 108n^3k^3 \ .
\end{aligned}$$

By applying Lemma 15 to $[\mathbf{Y}(E,F) - \alpha]$ we now get the following bound on the expectation

$$\begin{aligned}
\mathbb{E}\left[\left|e^{i\theta(\mathbf{Y}(E,F)-\alpha)} - 1\right|\right] &\leqslant \theta\sqrt{\mathrm{Var}\left(\mathbf{Y}(E,F) - \alpha\right)} \\
&\leqslant 6\sqrt{3}\theta n^{\frac{3}{2}}k^{\frac{3}{2}} \ . \qquad (2.49)
\end{aligned}$$

2.7.3.3 Bounding $\mathbb{E}\left[\left|e^{i\theta(\mathbf{Z}(E,F)-\beta)}-1\right|\right]$

To bound this, we use the same technique and setup as in the last subsection to show $\mathbf{Z}(E,F)$ is close to its expectation $\beta = \mathbb{E}[\mathbf{Z}(E,F)]$ with high probability. We do this by bounding the variance of $\mathbf{Z}(E,F)$, where Γ_3 denotes the set of all K_4 in K_n that have at least three edges in E.

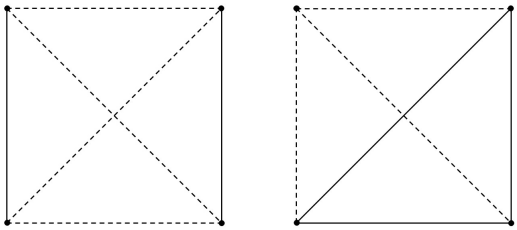


Figure 2.2: The only possible types of K_4 in Γ_3

In order to determine $|\Gamma_3|$, as well as the number of $r, s \in \Gamma_3$ such that $r \sim s$, we again count the number of r in Γ_3 that an edge $e \in E$ or $f \in F$ can be contained in. Suppose $e \in E$ is contained in $r \in \Gamma_3$. Then e must be adjacent to at least two other edges in E in $r \in \Gamma_3$. There are at most $3k^2$ choices for these edges. Note that since E is a collection of $K_{k,k}$ and hence triangle free, these three edges can not form a triangle and as a consequence fix all four vertices of $r \in \Gamma_3$.

Now suppose $f \in F$ is contained in $r \in \Gamma_3$. There are only two possible combinations of edges of E and F forming a K_4 in Γ_3 and these are presented in Figure 2.2. By inspection, note that $f \in F$ must be adjacent to at least one edge $e \in E$ in $r \in \Gamma_3$. There are at most $2k$ choices for this. In each case, every edge in E is adjacent to two other edges inside E and E does not contain any triangles. We can thus find another edge inside E in $r \in \Gamma_3$ adjacent to e which does not form a triangle with e and f and hence determines r completely. There are at most $2k$ choices for this. Thus, in total, f is contained in at most $4k^2$ copies $r \in \Gamma_3$.

In conclusion, for a given $s \in \Gamma_3$, there are at most $4 \cdot 3k^2 + 3 \cdot 4k^2 = 24k^2$ copies $s \in \Gamma_3$ such that $r \sim s$. Since there are at most nk edges in E and each of these is contained in at most k^2 copies $r \in \Gamma_3$, we have that $|\Gamma_3| \leqslant 3k^3 n$.

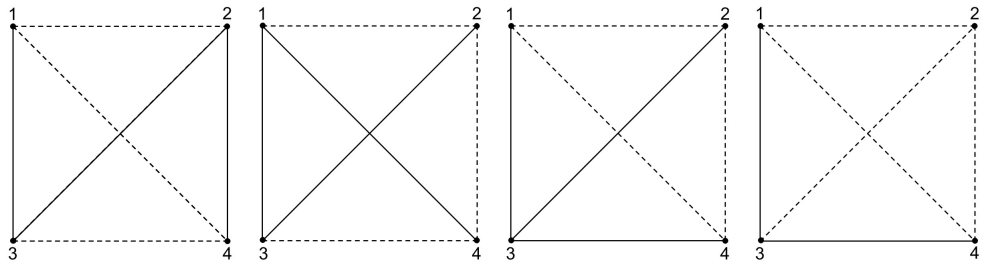


Figure 2.3: These types of K_4 are not possible in Γ_3

Thus

$$\begin{aligned}
\mathrm{Var}(\mathbf{Z}(E,F)) &= \sum_{r\in\Gamma}\sum_{s\in\Gamma}\mathrm{Cov}_{E,F}(T_r(E,F),T_s(E,F)) \\
&= \sum_{r\in\Gamma}\sum_{s\sim r}\mathrm{Cov}_{E,F}(T_r(E,F),T_s(E,F)) \text{ by independence} \\
&\leqslant 3nk^3\cdot 24k^2 \\
&\leqslant 72nk^5\ .
\end{aligned}$$

Thus applying Lemma 15 to $(\mathbf{Z}(E,F)-\beta)$ we get the following bound on the expectation

$$\begin{aligned}
\mathbb{E}\left[\left|e^{i\theta(\mathbf{Z}(E,F)-\beta)}-1\right|\right] &\leqslant \theta\cdot\sqrt{\mathrm{Var}\left(\mathbf{Z}(E,F)\right)} \\
&= O\left(\theta n^{1/2}k^{5/2}\right)\ . \qquad (2.50)
\end{aligned}$$

2.7.3.4 Combining Bounds

Combining the bounds on the expectations in equations (2.48), (2.49) and (2.50), we obtain the following bound on the absolute value of the expectation of the characteristic function

$$\begin{aligned}
\left|\mathbb{E}\left[e^{i\theta\mathbf{S}_n}\right]\right| &= \left|\mathbb{E}\left[e^{i\theta(\mathbf{C}(F)+\mathbf{W}(E,F))}\right]\right| + \mathbb{E}\left[\left|e^{i\theta(\mathbf{Y}(E,F)-\alpha)}-1\right|\right] + \mathbb{E}\left[\left|e^{i\theta(\mathbf{Z}(E,F)-\beta)}-1\right|\right] \text{ by (2.43)} \\
&= nke^{-\Theta(n)} + e^{-\Theta\left(\theta^2n^5k\right)} + O\left(\theta n^{3/2}k^{3/2}\right) + O\left(\theta n^{1/2}k^{5/2}\right) \\
&= nke^{-\Theta(n)} + e^{-\Theta\left(\theta^2n^5k\right)} + O\left(\theta n^{3/2}k^{3/2}\right)\ .
\end{aligned}$$

Now choose $k = \lfloor \frac{1}{3}\theta^{-\frac{5}{3}}n^{-4}\rfloor$ and $\delta = 0.5$. We claim that these choices work for all θ such that $\omega\left(n^{-3}\right) = \theta = o\left(n^{-2.4}\right)$. This holds as long as the following two conditions hold:

1. for all $\theta \in R_2$, so all θ such that $A/\theta_n = \theta = o\left(n^{-2.4}\right)$, equation (2.51) is bounded above by $(\theta\sigma_n)^{-1.5}$

2. k is sensible, so satisfies $1 \leqslant k = o(n)$

In order for condition 1 to hold, note that it is sufficient if each term of the sum is bounded above by $1/3(\theta\sigma_n)^{-1.5}$, which holds as long as

$$\omega\left(\theta^{-2}n^{-5}\log(n)\right) = k \leqslant \frac{1}{3}\theta^{-\frac{5}{3}}n^{-4}\ .$$

Our choice of $k = \lfloor \frac{1}{3}\theta^{-\frac{5}{3}}n^{-4} \rfloor$ satisfies this, so condition 1 holds.

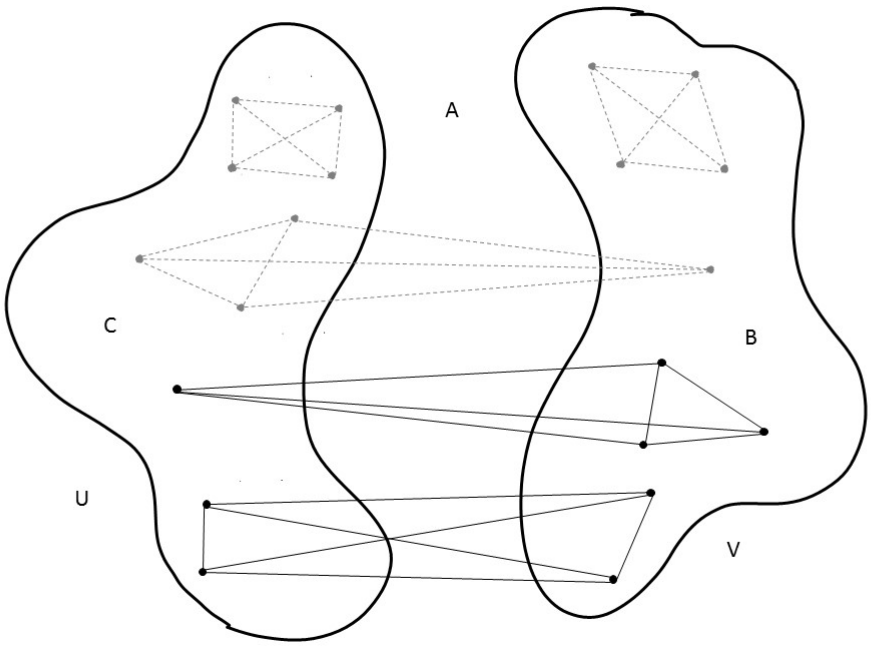
To establish condition 2, note that for $\omega\left(n^{-3}\right) = \theta = o\left(n^{-2.4}\right)$, we have that

$$\left\lfloor \frac{1}{3}\theta^{-\frac{5}{3}}n^{-4} \right\rfloor = o\left(n^5 n^{-4}\right) = o(n)$$
$$\left\lfloor \frac{1}{3}\theta^{-\frac{5}{3}}n^{-4} \right\rfloor = \omega\left((A/\sigma_n)^{-\frac{5}{3}}n^{-4}\right) \geqslant 1 \text{ for } A \text{ sufficiently large.}$$

Thus condition 2 also holds for our choice of k. This proves Lemma 8.

2.7.4 Region $R_5 = \left(n^{-1.96}, n^{-1.4}\right)$

We now prove Lemma 11. To bound the characteristic function in this region, we partition the vertex set $[n]$ into two parts, U and V, whose sizes we determine later. We denote by C all edges inside U, by B all edges inside V, and by A all edges between U and V.



For $u \in U$, let $\mathbf{I}_u^{(1)}(A)$ denote the vector of size $|V|$ that has entry i equal to one if the vertex i is connected to u in G and equal to zero otherwise. Similarly $\mathbf{I}_u^{(2)}(A)$ and $\mathbf{I}_u^{(3)}(A)$ are the vectors of size $\binom{|V|}{2}$ and $\binom{|V|}{3}$ where each entry corresponds to a set of two or three vertices inside V respectively rather than just a single vertex. The entry corresponding to $\{i,j\}$ or $\{i,j,k\}$ is equal to one if and only if $\{i,j\}$ or $\{i,j,k\}$ respectively are all

connected to the vertex u in G. For $u, w \in U$, we define the vector $\mathbf{I}^{(2)}_{u,w}(A)$ to be the vector of size $\binom{|V|}{2}$ in which every entry corresponds to a set of two vertices inside V and the entry corresponding to $\{i, j\}$ is equal to one if and only if both vertices i, j are connected to both u and w in G. Finally, the vector $\mathbf{M}^{(3)}(B)$ is the vector of size $\binom{|V|}{3}$ in which each entry corresponds to a set of three vertices inside V and this entry is equal to one if and only if all edges between these vertices exist and zero otherwise. The vector $\mathbf{M}^{(4)}(B)$ is defined analogously for sets of vertices of size four. Further, define $\triangle(C)$ to be the set of all triangles present between the vertices in U, and $E(C)$ the set of all edges present between the vertices in U.

Considering copies of K_4 with 0, 1, 2, 3 or 4 vertices in U, we see that the total number of K_4 in G can be written as

$$\begin{aligned}\mathbf{S}_n = &\left\langle \mathbf{M}^{(4)}(B), \mathbf{I} \right\rangle + \sum_{u \in U} \left\langle \mathbf{I}^{(3)}_u(A),\, \mathbf{M}^{(3)}(B) \right\rangle + \sum_{\{u,w\} \in E(C)} \left\langle \mathbf{I}^{(2)}_{u,w}(A),\, \mathbf{M}^{(2)}(B) \right\rangle \\ &+ \sum_{\{u,w,x\} \in \triangle(C)} \left\langle \mathbf{I}^{(1)}_{u,w,x}(A),\, \mathbf{I} \right\rangle + \left\langle \mathbf{M}^{(4)}(C), \mathbf{I} \right\rangle \ ,\end{aligned}$$

where $\mathbf{I}$ is the vector with ones in each entry. By applying Lemma 18 to $\mathbf{S}_n$ we get that

$$\left| \mathbb{E}\left[e^{i\theta \mathbf{S}_n} \right] \right|^2 \leqslant \mathbb{E}_{A,A',C}\left[\, |\mathbf{\Psi}| \, \right] \ , \tag{2.51}$$

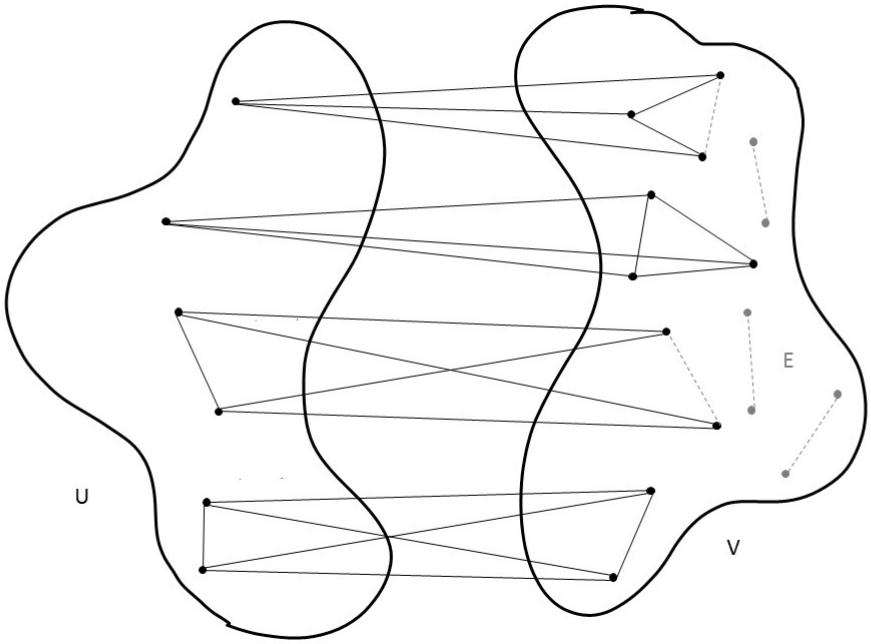
where

$$\mathbf{\Psi} = \mathbb{E}_B\left[e^{i\theta \mathbf{\Phi}} \right]$$

and

$$\mathbf{\Phi} = \sum_{u \in U} \left\langle \left(\mathbf{I}^{(3)}_u(A) - \mathbf{I}^{(3)}_u(A') \right), \mathbf{M}^{(3)}(B) \right\rangle + \sum_{\{u,w\} \in E(C)} \left\langle \left(\mathbf{I}^{(2)}_{u,w}(A) - \mathbf{I}^{(2)}_{u,w}(A') \right), \mathbf{M}^{(2)}(B) \right\rangle \ . \tag{2.52}$$

We aim to show that the expectation $\mathbb{E}\left[e^{i\theta \mathbf{S}_n} \right]$ is small in absolute value. We do so by showing that for most choices of C, A and A', the expectation $\mathbf{\Psi}$ is small in absolute value. In order to achieve this, we first condition on the edges inside A, A' and C, so treat these edges as fixed.



Let E be the set of all $\frac{|V|}{2}$ edges in a matching of V and let F be all edges in V outside E, so $B = E \cup F$. We will later also condition on all the edges outside E, so take these edges as fixed. We define $\mathbf{W}(A, A', C, F)$ to be the function equal to the all the terms in equation (2.52) that do not depend on edges in E, so the function $\mathbf{W}(A, A', C, F)$ satisfies

$$\begin{aligned}\mathbf{\Phi} = \mathbf{W}(A, A', C, F) &+ \sum_{e \in E} \sum_{u \in U} \sum_{v \in V} \left(\mathbf{I}_u^{(3)}(A) - \mathbf{I}_u^{(3)}(A')\right)_{e,v} \cdot \left(\mathbf{I}_v^{(2)}(F)\right)_e \cdot \mathbf{I}_e \\ &+ \sum_{e \in E} \sum_{\{u,w\} \in E(C)} \left(\mathbf{I}_{u,w}^{(2)}(A) - \mathbf{I}_{u,w}^{(2)}(A')\right)_e \cdot \mathbf{I}_e \ ,\end{aligned}$$

where $(.)_e$ denotes the entry corresponding to the edge e in a matrix and $\mathbf{I}_e$ is the indicator random variable of the presence of the edge e in G. Now define

$$\begin{aligned}\mathbf{Y}_e &= \sum_{u \in U} \sum_{v \in V} \left(\mathbf{I}_u^{(3)}(A) - \mathbf{I}_u^{(3)}(A')\right)_{e,v} \cdot \left(\mathbf{I}_v^{(2)}(F)\right)_e \\ \mathbf{Z}_e &= \sum_{\{u,w\} \in E(C)} \left(\mathbf{I}_{u,w}^{(2)}(A) - \mathbf{I}_{u,w}^{(2)}(A')\right)_e\end{aligned}$$

and note that neither $\mathbf{Y}_e$ nor $\mathbf{Z}_e$ depend on any of the edges inside E. We can thus rewrite $\mathbf{\Phi}$ as

$$\mathbf{\Phi} = \mathbf{W}(A, A', C, F) + \sum_{e \in E} (\mathbf{Y}_e + \mathbf{Z}_e) \cdot \mathbf{I}_e \tag{2.53}$$

and note that none of the terms in $\mathbf{\Phi}$ involve more than one edge in E. Substituting this

into equation (2.52)

$$\begin{aligned}
|\mathbf{\Psi}| &= \left|\mathbb{E}_{E,F}\left[e^{i\theta\cdot\left(\mathbf{W}(A,A',C,F)+\sum_{e\in E}(\mathbf{Y}_e+\mathbf{Z}_e)\cdot\mathbf{I}_e\right)}\right]\right| \\
&\leqslant \mathbb{E}_F\left[\left|\mathbb{E}_E\left[e^{i\theta\cdot\left(\mathbf{W}(A,A',C,F)+\sum_{e\in E}(\mathbf{Y}_e+\mathbf{Z}_e)\cdot\mathbf{I}_e\right)}\right]\right|\right] \\
&= \mathbb{E}_F\left[\left|e^{i\theta\cdot\mathbf{W}(A,A',C,F)}\cdot\mathbb{E}_E\left[e^{i\theta\sum_{e\in E}(\mathbf{Y}_e+\mathbf{Z}_e)\cdot\mathbf{I}_e}\right]\right|\right] \\
&= \mathbb{E}_F\left[\left|\mathbb{E}_E\left[e^{i\theta\cdot\sum_{e\in E}(\mathbf{Y}_e+\mathbf{Z}_e)\cdot\mathbf{I}_e}\right]\right|\right] \ . \qquad (2.54)
\end{aligned}$$

Thus the problem of bounding $|\mathbf{\Psi}|$ reduces to showing that for most choices of F the expectation $\mathbf{\Lambda} = \mathbb{E}_E\left[e^{i\theta\sum_{e\in E}(\mathbf{Y}_e+\mathbf{Z}_e)\cdot\mathbf{I}_e}\right]$ is small in absolute value. We will later condition on all edges outside E. After conditioning on these, $\mathbf{\Phi}$ is just a constant plus the sum of $|E|$ independent Bernoulli random variables each multiplied by a constant. So as long as we can show that the constants we multiply the Bernoulli random variables by are typically of the same size, around $|U|^{0.5}|V|$ for the range of θ we are aiming for in this section, we can use this to bound the expectation of the characteristic function in a similar fashion as outlined in Section 2.6. In the following we will denote an observation of $\mathbf{Y}_e$ by $\mathbf{y}_e$ and an observation of $\mathbf{Z}_e$ by $\mathbf{z}_e$.

To show that the value of $|\mathbf{\Lambda}|$ is small, we show that with high probability a certain 'good' event $\mathfrak{X}$ involving the edges A, A', C and F occurs, and that conditional on this event, the absolute value of $e^{i\theta\sum_{e\in E}(\mathbf{Y}_e+\mathbf{Z}_e)\cdot\mathbf{I}_e}$ is small. For an edge $e \in E$, we let $\mathfrak{X}_e$ be the event that the random variables $\mathbf{Y}_e$ and $\mathbf{Z}_e$ are both of roughly the size we expect them to be. The event $\mathfrak{X}$ is then defined to be the event that $\mathfrak{X}_e$ holds for at least $\frac{4}{10}$ of all edges $e \in E$. We wish to show, that the event $\mathfrak{X}$ holds with high probability. To do so, we consider three good events $\boldsymbol{x}_e$, $\boldsymbol{y}_e$, $\boldsymbol{z}_e$ on each edge $e \in E$. We then show that for any fixed edge $e \in E$, the probability $\mathbb{P}(\boldsymbol{x}_e) \geq 0.45$ and both $\boldsymbol{y}_e$ and $\boldsymbol{z}_e$ hold with exponentially small failure probability. Denoting by T the set of all edges for which $\boldsymbol{x}_e$ holds, we can thus show that with high probability $|T| \geq \frac{4}{10}|E|$ and both $\boldsymbol{y}_e$ and $\boldsymbol{z}_e$ also hold for all edges inside T. We further show that conditional on $\boldsymbol{x}_e$ and $\boldsymbol{y}_e$, $\mathbf{Y}_e$ is with high probability of the right size. Similarly, we show that conditional on $\boldsymbol{x}_e$ and $\boldsymbol{z}_e$, $\mathbf{Z}_e$ is with high probabiliy of the right size. This implies that with high probability for all $e \in T$, $\mathbf{Y}_e$ and $\mathbf{Z}_e$ are both of the right size and $|T| \geq \frac{4}{10}|E|$, which gives us the required result that $\mathfrak{X}$ occurs with high probability.

We will now go on to define these events. For an edge $e \in E$, we let $\mathcal{Y}_e$ be the event that $\mathbf{Y}_e$ is of roughly the right size, so in particular the event that

$$\mathbf{Y}_e \in \left[\frac{p^3}{4}|U|^{0.49}|V|, \frac{9p^3}{4}|U|^{0.51}|V|\right] \ .$$

Similarly, we define the event $\mathscr{E}_e$ to be the event that $\mathbf{Z}_e$ is roughly of the right size, so in particular the event that

$$\mathbf{Z}_e \in \left[\frac{p^3}{8}|U|^{1.49}, \frac{9p^3}{4}|U|^{1.51}\right] .$$

The event $\mathscr{x}_e$, which we will show to hold with high probability for at least $\frac{4}{10}$ of all edges $e \in E$, is the event that the difference between the number of common neighbours of the endvertices of e between A and A' is roughly of size $\sqrt{|U|}$, which is how large we expect it to be on average. Formally, we define $\mathscr{x}_e$ to be the event that

$$\mathscr{x}_e := \left\{\sum_{u\in U}(\mathbf{I}_u^{(2)}(A) - \mathbf{I}_u^{(2)}(A'))_e \in [|U|^{0.49}, |U|^{0.51}]\right\} .$$

Let further

$$T = \{e \in E \mid \mathscr{x}_e \text{ occurs}\} . \tag{2.55}$$

and define $\mathscr{x}$ to be the event that $|T| \geq \frac{4}{10}|E|$. The only two events left to be defined are the events $\mathscr{y}_e$ and $\mathscr{z}_e$. For a fixed edge $e \in E$, let V_e denote the set of common neighbours of the endvertices of e that lie in V, and let m_e denote the size of V_e. We then define $\mathscr{y}_e$ to be the event that m_e is of roughly the size we expect it to be, so $\mathscr{y}_e$ is the event that

$$\mathscr{y}_e = \left\{m_e \in \left[\frac{p^2|V|}{2}, \frac{3p^2|V|}{2}\right]\right\}$$

and let $\mathscr{y}$ be the event that $\mathscr{y}_e$ holds for all $e \in E$.

Note that since E is a matching, the events $\mathscr{y}_e$ only depend on edges in F. For an edge $e \in E$, let S_e denote the set of vertices in U that are connected to both endvertices of e in A. Similarly, let S'_e denote the set of vertices in U that are connected to both endvertices of e in A'. For a set M of vertices, we further denote by $\mathscr{C}(M)$ be the number of connected pairs of vertices in M. For an edge $e \in E$, we let $\mathscr{z}_e$ be the event that the number of common neighbours of the endvertices of e in U is of roughly the size we expect it to be, more precisely, $\mathscr{z}_e$ is the event that

$$\mathscr{z}_e = \left\{|S_e| \in \left[\frac{p^2}{2}|U|, \frac{3p^2}{2}|U|\right]\right\} . \tag{2.56}$$

Furthermore, $\mathscr{z}$ is the event that $\mathscr{z}_e$ holds for all $e \in E$.

Having defined these events, we now go on to bound the probability of $\mathscr{X}$ occurring. In the following we assume that $|U|$ and $|V|$ are both of size at least $\sqrt{n}$.

We start by bounding the probability of the event $\mathscr{x}$.

Lemma 30. *We have that* $\mathbb{P}(\mathscr{x}) \geq 1 - e^{-\Theta(|V|)}$.

Proof. Note that $\sum_{u\in U}(\mathbf{I}_u^{(2)}(A)-\mathbf{I}_u^{(2)}(A'))_e$ is the sum of $|U|$ random variables, each of which is equal to 1 with probability $p^2(1-p^2)$, 0 with probability $1-2(p^2-p^4)$ and equal to -1 with probability $p^2(1-p^2)$ independently of all others. By standard concentration and anti-concentration estimates for binomial distributions, we have that $\mathbb{P}(\boldsymbol{x}_e)\geq 0.45$, as long as n is sufficiently large, depending on p. Since the edges of E form a matching, the events $\boldsymbol{x}_e$ are independent. The number of edges $e\in E$ for which $\boldsymbol{x}_e$ occurs is hence Binomially distributed with parameters $|E|=|V|/2$ and $\mathbb{P}(\boldsymbol{x}_e)$. Thus, by using the Chernoff bound from Theorem 21, $\boldsymbol{x}_e$ holds for at least $\frac{4}{10}$ of all edges in the matching with a failure probability that is exponentially small in $|V|$ as required. ■

Lemma 31. *We have that* $\mathbb{P}(\boldsymbol{y})\geq 1-|V|e^{-\Theta(|V|)}$.

Proof. Note that for a fixed edge $e\in E$ the random variable m_e is binomially distributed with parameters $|V|-2$ and p^2. By applying the Chernoff bound from Theorem 21 we thus have that e has at least $\frac{p^2|V|}{2}$ and at most $\frac{3p^2|V|}{2}$ common neighbours in V with probability at least $1-e^{-\Theta(|V|)}$. By a union bound applied to the complements, $\boldsymbol{y}_e$ holds for all $e\in E$ with probability at least $1-|E|e^{-\Theta(|V|)}$. Thus

$$\mathbb{P}(\boldsymbol{y})\geq 1-|V|e^{-\Theta(|V|)}\ ,$$

as required. ■

Lemma 32. *We have that* $\mathbb{P}(\boldsymbol{z})\geq 1-|V|e^{-\Theta(|U|)}$.

Proof. For an edge $e\in E$, note that the number of common neighbours in U is binomially distributed with parameters $|U|$ and p^2. Thus by applying the Chernoff bound from Theorem 21 we have that

$$\mathbb{P}(\boldsymbol{z}_e)=\mathbb{P}\left(\frac{p^2}{2}|U|\leqslant\sum_{u\in U}X_u^e\leqslant\frac{3p^2}{2}|U|\right)\geq 1-e^{-\Theta(|U|)}\ . \tag{2.57}$$

Now applying a union bound to $\boldsymbol{z}_e^{\mathrm{c}}$ for all $e\in E$,

$$\mathbb{P}(\boldsymbol{z})\geq 1-|E|e^{-\Theta(|U|)}\geq 1-|V|e^{-\Theta(|U|)}\ ,$$

as required. ■

From now on we condition on the events $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}$ occurring. Conditional on $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}$ occurring we now wish to show that for at least $\frac{4}{10}$ of all edges in E the random variables $\mathbf{Y}_e$ and $\mathbf{Z}_e$ are with high probability of roughly the size we expect them to be. Recall that we denote by T the set of edges $e\in E$ for which $\boldsymbol{x}_e$ holds.

Lemma 33. *For a fixed edge* $e\in E$, *we have that* $\mathbb{P}\left(\boldsymbol{Y}_e^{\mathrm{c}}\mid\boldsymbol{x}_e,\boldsymbol{y}_e\right)\leqslant e^{-\Theta(n^{0.48})}$.

Proof. Fix an edge $e \in E$. We now condition on all edges between e and U and all edges inside U and F. This determines whether $\boldsymbol{y}_e$ and $\boldsymbol{x}_e$ occur. Throughout the rest of the proof, we will condition on these two events occurring. Recall that for an edge e, we denote by V_e the set of common neighbours in V of the endvertices of e. For a given vertex $u \in U$ the following four possible events can occur

- u is connected to both endvertices of e in A but not in A', denoted by $\boldsymbol{+}_{e,u}$
- u is connected to both endvertices of e in A' but not in A, denoted by $\boldsymbol{-}_{e,u}$
- u is connected to both endvertices of e in both A and A', denoted by $\boldsymbol{\pm}_{e,u}$
- u is not connected to both endvertices of e in neither A nor A', denoted by $\mathbf{0}_{e,u}$

We let $\gamma_{e,u,v}$ be the indicator random variable for the presence of the K_4^- formed by the endvertices of e, u and v in A minus the indicator random variable of the presence of the K_4^- formed by the endvertices of e, u and v in A', so

$$\mathbf{Y}_e = \sum_{u,v} \gamma_{e,u,v} \ . \tag{2.58}$$

If we condition on all edges in U and F and between e and U, this determines which of the events $\boldsymbol{+}_{e,u}$, $\boldsymbol{-}_{e,u}$, $\boldsymbol{\pm}_{e,u}$ and $\mathbf{0}_{e,u}$ occurs. After the conditioning, the only remaining randomness for $\gamma_{e,u,v}$ is the presence or absence of the edge uv in A and A'. Looking at the above defined four possible events we have

- if the event $\boldsymbol{+}_{e,u}$ occurs and $v \in V_e$

$$\begin{aligned} \mathbb{P}(\gamma_{e,u,v} = 1) &= p \\ \mathbb{P}(\gamma_{e,u,v} = -1) &= 0 \\ \mathbb{P}(\gamma_{e,u,v} = 0) &= 1 - p \end{aligned}$$

- if the event $\boldsymbol{-}_{e,u}$ occurs and $v \in V_e$

$$\begin{aligned} \mathbb{P}(\gamma_{e,u,v} = 1) &= 0 \\ \mathbb{P}(\gamma_{e,u,v} = -1) &= p \\ \mathbb{P}(\gamma_{e,u,v} = 0) &= 1 - p \end{aligned}$$

- if the event $\boldsymbol{\pm}_{e,u}$ occurs and $v \in V_e$

$$\begin{aligned} \mathbb{P}(\gamma_{e,u,v} = 1) &= p(1-p) \\ \mathbb{P}(\gamma_{e,u,v} = -1) &= p(1-p) \\ \mathbb{P}(\gamma_{e,u,v} = 0) &= 1 - 2(p(1-p)) \end{aligned}$$

- if the event $\mathbf{0}_{e,u}$ occurs and $v \in V_e$

$$\mathbb{P}(\gamma_{e,u,v} = 0) = 1 \ .$$

Furthermore, for a fixed edge e and conditional on all edges between U and e, the random variables $\{\gamma_{e,u,v}\}_{u\in U, v\in V}$ are independent. Let $\mathscr{F}$ denote the σ-algebra generated by everything we have conditioned on thus far.

For an edge $e \in E$, we denote by $\widetilde{\mathbf{Y}}_e$ the random variable $\mathbf{Y}_e$ conditional on $\mathscr{F}$. Furthermore, letting $\mathbf{H}_e = \sum_{u\in U} \left(\mathbf{I}_u^{(2)}(A) - \mathbf{I}_u^{(2)}(A')\right)_e$, we note that

$$\mathbb{E}\left[\widetilde{\mathbf{Y}}_e\right] = p\mathbf{H}_e m_e \tag{2.59}$$

and since we are assuming that the event $\boldsymbol{x}_e$ holds for the edge we are considering, we have that

$$\mathbf{H}_e \in \left[|U|^{0.49}, |U|^{0.51}\right] \ . \tag{2.60}$$

Since we have further conditioned on the event $\boldsymbol{y}_e$ occurring, we have

$$m_e \in \left[\frac{p^2|V|}{2}, \frac{3p^2|V|}{2}\right] \ . \tag{2.61}$$

Substituting equations (2.60) and (2.61) into equation (2.59) we have the following bound on the expectation of $\widetilde{\mathbf{Y}}_e$

$$\frac{p^3}{2}|U|^{0.49}|V| \leqslant \mathbb{E}\left[\widetilde{\mathbf{Y}}_e\right] \leqslant \frac{3p^3}{2}|U|^{0.51}|V| \ .$$

We now apply the Chernoff bound from Theorem 22, which is possible as $\widetilde{\mathbf{Y}}_e$ is the sum of $|U| \cdot m_e$ random variables satisfying the conditions of Theorem 22. Recall that $\mathbb{E}\left[\widetilde{\mathbf{Y}}_e\right] = p\, m_e\, \mathbf{H}_e$. So

$$\begin{aligned}
\mathbb{P}\left(\left|\widetilde{\mathbf{Y}}_e - \mathbb{E}\left[\widetilde{\mathbf{Y}}_e\right]\right| \geq \frac{1}{2}\mathbb{E}\left[\widetilde{\mathbf{Y}}_e\right]\right) &\leqslant 2 \cdot \exp\left\{-\frac{\frac{1}{2}\mathbb{E}\left[\widetilde{\mathbf{Y}}_e\right]^2}{|U| \cdot m_e}\right\} \\
&\leqslant e^{-\Theta(m_e \mathbf{H}_e^2/|U|)} \\
&= e^{-|U|^{-0.02}|V|} \text{ since the events } \boldsymbol{x}_e \text{ and } \boldsymbol{y}_e \text{ occur} \\
&= e^{-\Theta(n^{0.48})} \text{ since } |U| \leqslant n \text{ and } |V| \geq \sqrt{n} \ .
\end{aligned}$$

So as long as $\boldsymbol{x}_e$ and $\boldsymbol{y}_e$ both hold, with probability at least $1 - e^{-\Theta(n^{0.48})}$

$$\mathbf{Y}_e \in \left[\frac{p^3}{4}|U|^{0.49}|V|, \frac{9p^3}{4}|U|^{0.51}|V|\right] \ .$$

Note that this is equivalent to the event $\mathscr{Y}_e$ occurring, so $\mathbb{P}(\mathscr{Y}_e \mid \boldsymbol{x}_e, \boldsymbol{y}_e) \geq 1 - e^{-\Theta(n^{0.48})}$ as required. ■

Using a union bound on the complements, we thus have that conditional on $\boldsymbol{x}$ and $\boldsymbol{y}$ with probability at least $1 - |V|e^{-\Theta(n^{0.48})}$ the event $\mathscr{Y}_e$ holds for all $e \in T$. We now need to show the equivalent result for $\mathbf{Z}_e$, so that for all $e \in T$, $\mathbf{Z}_e$ is also of roughly the size we expect it to be. Recall that $\mathscr{Z}_e$ is the event that $\mathbf{Z}_e \in [p^3/8|U|^{1.49}, 9p^3/4|U|^{1.51}]$. We now bound the probability of $\mathscr{Z}_e$ occurring conditional on the events $\boldsymbol{z}_e$ and $\boldsymbol{x}_e$.

Lemma 34. *For any fixed edge $e \in E$, we have that $\mathbb{P}(\mathscr{Z}_e^{\mathrm{c}} \mid \boldsymbol{z}_e, \boldsymbol{x}_e) \leqslant e^{-\Theta\left(|U|^{0.98}\right)}$.*

Proof. Fix an edge $e \in E$ and condition on all edges outside E and U. This determines whether $\boldsymbol{x}_e$ and $\boldsymbol{z}_e$ occur. Assume they do both occur. We denote by $\mathscr{F}$ the σ-algebra generated by everything we have conditioned on thus far, so all edges outside E, and denote by $\widetilde{\mathbf{Z}}_e$ the random variable equal to the random variable $\mathbf{Z}_e$ conditional on $\mathscr{F}$. Note that

$$\widetilde{\mathbf{Z}}_e = \mathscr{C}(S_e) - \mathscr{C}(S'_e) \ .$$

Since we are assuming that $\boldsymbol{z}_e$ occurs, we know that S_e satisfies

$$|S_e| \in \left[\frac{p^2}{2}|U|, \frac{3p^2}{2}|U|\right] \ . \tag{2.62}$$

We are also assuming that $\boldsymbol{x}_e$ occurs, which means that

$$\sum_{u \in U} \left(\mathbf{I}_u^{(2)}(A) - \mathbf{I}_u^{(2)}(A')\right)_e = |S_e| - |S'_e| \in \left[|U|^{0.49}, |U|^{0.51}\right] \ .$$

This implies that

$$|S'_e| \in \left[|S_e| - |U|^{0.51}, |S_e| - |U|^{0.49}\right] \ . \tag{2.63}$$

The expectation of $\widetilde{\mathbf{Z}}_e$ is

$$\mathbb{E}\left[\widetilde{\mathbf{Z}}_e\right] = p\binom{|S_e|}{2} - p\binom{|S'_e|}{2} \ .$$

As we are assuming that both $\boldsymbol{x}_e$ and $\boldsymbol{z}_e$ occur, it follows that

$$\begin{aligned}
\mathbb{E}\left[\widetilde{\mathbf{Z}}_e\right] &= p\binom{|S_e|}{2} - p\binom{|S'_e|}{2} \\
&\geq p\binom{|S_e|}{2} - p\binom{|S_e| - |U|^{0.49}}{2} \text{ by (2.63)} \\
&= \frac{p}{2}\left(|S_e|^2 - |S_e|\right) - \frac{p}{2}\left(|S_e|^2 - 2|S_e||U|^{0.49} - |S_e| + |U|^{0.98} - |U|^{0.49}\right) \\
&= \frac{p}{2}\left(2|S_e||U|^{0.49} - |U|^{0.98} + |U|^{0.49}\right) \\
&\geq \frac{p}{2}\left(p^2|U|^{1.49} - |U|^{0.98} + |U|^{0.49}\right) \text{ by (2.62)} \\
&\geq \frac{p^3}{4}|U|^{1.49} \ ,
\end{aligned}$$

for n large enough. Similarly

$$\begin{aligned}
\mathbb{E}\left[\widetilde{\mathbf{Z}}_e\right] &= p\binom{|S_e|}{2} - p\binom{|S'_e|}{2} \\
&\leqslant p\binom{|S_e|}{2} - p\binom{|S_e| - |U|^{0.51}}{2} \text{ by (2.63)} \\
&= \frac{p}{2}\left(|S_e|^2 - |S_e|\right) - \frac{p}{2}\left(|S_e|^2 - 2|S_e||U|^{0.51} - |S_e| + |U|^{1.02} - |U|^{0.51}\right) \\
&= \frac{p}{2}\left(2|S_e||U|^{0.51} - |U|^{1.02} + |U|^{0.51}\right) \\
&\leqslant \frac{p}{2}\left(3p^2|U|^{1.51} - |U|^{1.02} + |U|^{0.51}\right) \text{ by (2.62)} \\
&\leqslant \frac{3p^3}{2}|U|^{1.51} \ .
\end{aligned}$$

It follows that for n large enough

$$\frac{p^3}{4}|U|^{1.49} \leqslant \mathbb{E}\left[\widetilde{\mathbf{Z}}_e\right] \leqslant \frac{3p^3}{2}|U|^{1.51} \ . \tag{2.64}$$

We now apply the Chernoff bound from Theorem 22 to $\widetilde{\mathbf{Z}}_e$, which is possible since $\mathscr{C}(S_e) = \sum_{0\leqslant i<j\leqslant|S_e|} X_{i,j}$ where $X_{i,j} = 1$ if the i^{th} and j^{th} vertex in S_e are connected by an edge and $X_{i,j} = 0$ otherwise and similarly for $\mathscr{C}(S'_e)$. So $\widetilde{\mathbf{Z}}_e = \mathscr{C}(S_e) - \mathscr{C}(S'_e)$ is the sum of at most $\binom{|S_e|+|S'_e|}{2} = O\left(|U|^2\right)$ random variables each with possible values in either $\{-1, 0\}$ or $\{0, 1\}$ and satisfying the conditions of Theorem 22. This implies

$$\begin{aligned}
\mathbb{P}\left(\left|\widetilde{\mathbf{Z}}_e - \mathbb{E}\left[\widetilde{\mathbf{Z}}_e\right]\right| \geq \frac{1}{2}\mathbb{E}\left[\widetilde{\mathbf{Z}}_e\right]\right) &\leqslant 2\exp\left\{-\frac{\frac{1}{2}\mathbb{E}\left[\widetilde{\mathbf{Z}}_e\right]^2}{\Theta\left(|U|^2\right)}\right\} \\
&= e^{-\Theta\left(|U|^{0.98}\right)} \ .
\end{aligned} \tag{2.65}$$

Note by also recalling equation (2.64), that the condition

$$\left|\widetilde{\mathbf{Z}}_e - \mathbb{E}\left[\widetilde{\mathbf{Z}}_e\right]\right| \geq \frac{1}{2}\mathbb{E}\left[\widetilde{\mathbf{Z}}_e\right] \tag{2.66}$$

is a stricter condition than $\mathscr{Z}_e^{\mathrm{c}}$ occurring, so whenever (2.66) holds, it is certainly satisfied that

$$\mathbf{Z}_e \in \left[\frac{p^3}{8}|U|^{1.49}, \frac{9p^3}{4}|U|^{1.51}\right] \ .$$

As a consequence, conditional on $\boldsymbol{x}_e$ and $\boldsymbol{\varkappa}_e$, the event $\mathscr{Z}_e^{\mathrm{c}}$ occurring is at least as likely as $\left|\widetilde{\mathbf{Z}}_e - \mathbb{E}\left[\widetilde{\mathbf{Z}}_e\right]\right| \geq \frac{1}{2}\mathbb{E}\left[\widetilde{\mathbf{Z}}_e\right]$ occurring. In other words

$$\mathbb{P}\left(\mathscr{Z}_e^{\mathrm{c}} \mid \boldsymbol{x}_e, \boldsymbol{\varkappa}_e\right) \leqslant e^{-\Theta\left(|U|^{0.98}\right)} \ ,$$

as required. ■

By applying a union bound over all $e \in T$, and so conditioning on all of $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}$ occurring, we have that with probability at least $1 - |V|e^{-\Theta\left(|U|^{0.98}\right)} - |V|e^{-\Theta(n^{0.48})}$ the events $\mathcal{X}_e$ and $\mathcal{Y}_e$ both hold for all $e \in T$ and further the size of T satisfies $|T| \geq \frac{4}{10}|E|$. Recall that this is equivalent to the event $\mathfrak{X}$ occurring. We can now finally bound the probability of $\mathfrak{X}$ occurring.

Lemma 35. *We have that* $\mathbb{P}(\mathfrak{X}) \geqslant 1 - e^{-\Theta(n^{0.47})}$.

Proof. For a fixed edge $e \in E$, by Lemma 33 and Lemma 34

$$\begin{aligned}\mathbb{P}\left(\mathfrak{X}_e^{\mathrm{c}} \cap \boldsymbol{x}_e \cap \boldsymbol{y}_e \cap \boldsymbol{z}_e\right) &\leqslant e^{-\Theta\left(|U|^{0.98}\right)} + e^{-\Theta\left(n^{0.48}\right)} \\ &\leqslant e^{-\Theta\left(n^{0.48}\right)} \ .\end{aligned}$$

Applying a union bound over all $e \in E$, we have

$$\mathbb{P}\left(\mathfrak{X}^{\mathrm{c}} \cap \boldsymbol{x} \cap \boldsymbol{y} \cap \boldsymbol{z}\right) \leqslant |V|e^{-\Theta\left(n^{0.48}\right)} \ . \tag{2.67}$$

Combining this with Lemmas 30, 31 and 32 we see that

$$\begin{aligned}\mathbb{P}\left(\mathfrak{X}^{\mathrm{c}}\right) &= \mathbb{P}\left(\mathfrak{X}^{\mathrm{c}} \cap \boldsymbol{x} \cap \boldsymbol{y} \cap \boldsymbol{z}\right) + \mathbb{P}\left(\mathfrak{X}^{\mathrm{c}} \cap \left(\boldsymbol{x}^{\mathrm{c}} \cup \boldsymbol{y}^{\mathrm{c}} \cup \boldsymbol{z}^{\mathrm{c}}\right)\right) \\ &\leqslant \mathbb{P}\left(\mathfrak{X}^{\mathrm{c}} \cap \boldsymbol{x} \cap \boldsymbol{y} \cap \boldsymbol{z}\right) + \mathbb{P}\left(\boldsymbol{x}^{\mathrm{c}} \cup \boldsymbol{y}^{\mathrm{c}} \cup \boldsymbol{z}^{\mathrm{c}}\right) \\ &\leqslant |V|e^{-\Theta\left(n^{0.48}\right)} + \left(e^{-\Theta(|V|)} + e^{-\Theta(|U|)}\right) \\ &\leqslant e^{-\Theta(n^{0.47})} \text{ since we assumed } |U|, |V| \geqslant \sqrt{n} \ ,\end{aligned}$$

completing the proof. ∎

Having shown that both $\mathbf{Y}_e$ and $\mathbf{Z}_e$ are with high probability contained in a small interval, we can finally use this to find a bound on $|\Lambda|$ when the event $\mathfrak{X}$ occurs. To cover a wider range of θ, we will consider two different cases for the size of U and V.

Lemma 36. *For all* $\theta \in (n^{-1.96}, n^{-1.4})$*, when* $\mathfrak{X}$ *holds, we have*

$$\left|\mathbb{E}_E\left[e^{i\theta\sum_{e\in E}(\mathbf{Y}_e+\mathbf{Z}_e)\cdot\mathbf{I}_e}\right]\right| \leqslant e^{-\Theta(n^{0.04})} \ .$$

Proof. We prove this in two steps, choosing different values for the size of U and V in each step. Recall that $\mathbb{E}_E\left[e^{i\theta\sum_{e\in E}(\mathbf{Y}_e+\mathbf{Z}_e)\cdot\mathbf{I}_e}\right]$ is a random variable depending only on the edges outside E. Condition on all the edges outside E. This determines whether the event $\mathfrak{X}$ occurs, so suppose it does. Letting $\mathbf{S}_E := \sum_{e\in E}(\mathbf{y}_e + \mathbf{z}_e)\cdot\mathbf{I}_e$, we see that $\mathbf{S}_E$ is the sum of $|E|$ independent identically distributed random Bernoulli random variables, each multiplied by a constant. Now consider the case that $|U| = \frac{n}{2}$ and $|V| = \frac{n}{2}$. This will give us the desired bound for the range $n^{-1.96} < \theta = o\left(n^{-1.51}\right)$. For all edges $e \in E$ for which the event $\mathfrak{X}_e$ occurs,

$$\mathbf{y}_e + \mathbf{z}_e \in \left[\frac{p^3}{8}n^{1.49},\ \frac{9p^3}{4}n^{1.51}\right] \ .$$

Since we have assumed that the event $\mathfrak{X}$ occurs, at least 4/10 of these constants, so at least $\frac{4}{10}|E| = \frac{n}{10}$, lie in the interval $\left[\frac{p^3}{8}n^{1.49}, \frac{9p^3}{4}n^{1.51}\right]$. This allows us to bound the conditional expectation as follows:

$$\begin{aligned}
\left|\mathbb{E}_E\left[e^{i\theta\mathbf{S}_E}\right]\right| &= \left|\mathbb{E}_E\left[\prod_{e\in E} e^{i\theta(\mathbf{y}_e+\mathbf{z}_e)\cdot\mathbf{I}_e}\right]\right| \\
&= \prod_{e\in E}\left|\mathbb{E}_E\left[e^{i\theta(\mathbf{y}_e+\mathbf{z}_e)\cdot\mathbf{I}_e}\right]\right| \text{ since the } \mathbf{I}_e \text{ are independent} \\
&\leqslant \prod_{e\in E}\left(1-8p(1-p)\left|\left|\frac{\theta\left(\mathbf{y}_e+\mathbf{z}_e\right)}{2\pi}\right|\right|^2\right) \\
&= \prod_{e\in E,\mathfrak{X}_e}\left(1-8p(1-p)\left(\frac{\theta\left|\mathbf{y}_e+\mathbf{z}_e\right|}{2\pi}\right)^2\right) \text{ as } \mathbf{y}_e+\mathbf{z}_e \leqslant \frac{9p^3}{4}n^{-1.51}, \theta = o\left(n^{1.51}\right) \\
&\leqslant \prod_{e\in E,\mathfrak{X}_e}\left(1-8p(1-p)\left(\frac{\theta\left(p^4\cdot n^{1.49}\right)}{16\pi}\right)^2\right) \text{ as } \mathbf{y}_e+\mathbf{z}_e \geq \frac{p^3n^{1.49}}{8} \\
&\leqslant \left(1-8p(1-p)\left(\frac{\theta\left(p^3n^{1.49}\right)}{16\pi}\right)^2\right)^{\frac{n}{10}} \\
&= e^{-\Omega(\theta^2n^{3.98})} .
\end{aligned} \tag{2.68}$$

Combining this with Lemma 35 we have

$$\begin{aligned}
\mathbb{E}\left[e^{i\theta\mathbf{S}_n}\right] &\leqslant \left(\mathbb{E}_{A,A',C}\left[|\mathbf{\Psi}|\right]\right)^{1/2} \text{ by (2.51)} \\
&\leqslant \left(\mathbb{E}\left[\left|\mathbb{E}_E\left[e^{i\theta\sum_{e\in E}(\mathbf{Y}_e+\mathbf{Z}_e)\cdot\mathbf{I}_e}\right]\right|\right]\right)^{1/2} \text{ by (2.54)} \\
&\leqslant \left(\mathbb{P}(\mathfrak{X}^{\mathrm{c}}) + \max_{\mathbf{y}_e+\mathbf{z}_e|\mathfrak{X} \text{ occurs}} \mathbb{E}\left[\left|\mathbb{E}_E\left[e^{i\theta\sum_{e\in E}(\mathbf{y}_e+\mathbf{z}_e)\cdot\mathbf{I}_e}\right]\right|\right]\right)^{1/2} \\
&\leqslant \left(e^{-\Theta(n^{0.47})} + e^{-\Omega(\theta^2\cdot n^{3.98})}\right)^{1/2} \text{ by Lemma 35 and (2.68)} \\
&\leqslant e^{-\Theta(n^{0.47})} + e^{-\Omega(\theta^2\cdot n^{3.98})} ,
\end{aligned} \tag{2.69}$$

which is at most $e^{-\Theta(n^{0.06})}$ for $n^{-1.96} < \theta = o\left(n^{-1.51}\right)$.

We now also consider the case when $|U| = \sqrt{n}$ and $|V| = n - \sqrt{n}$. This will allow us to cover the range $\omega\left(n^{-1.6}\right) < \theta < n^{-1.4}$. In this case we have for all edges for which $\mathfrak{X}_e$ occurs, by the definition of $\mathfrak{X}_e$ that

$$\mathbf{y}_e + \mathbf{z}_e \in \left[\frac{p^3}{4}n^{1.245}, \frac{5p^3}{2}n^{1.255}\right] .$$

Since we have assumed that the event $\mathfrak{X}$ occurs, for at least 4/10 of the edges $e \in E$, $\mathbf{y}_e + \mathbf{z}_e$ lies in this interval. This allows us to bound the conditional expectation as follows:

$$\begin{aligned}
\left|\mathbb{E}_E\left[e^{i\theta \mathbf{S}_E}\right]\right| &= \left|\mathbb{E}_E\left[e^{i\theta \sum (\mathbf{Y}_e + \mathbf{Z}_e)\cdot \mathbf{I}_e}\right]\right| \\
&= \left|\mathbb{E}_E\left[\prod_{e\in E} e^{i\theta(\mathbf{y}_e+\mathbf{z}_e)\cdot \mathbf{I}_e}\right]\right| \\
&= \prod_{e\in E}\left|\mathbb{E}_E\left[e^{i\theta(\mathbf{y}_e+\mathbf{z}_e)\cdot \mathbf{I}_e}\right]\right| \text{ since } \mathbf{I}_e \text{ are independent} \\
&\leqslant \prod_{e\in E}\left(1 - 8p(1-p)\left|\left|\frac{\theta\left(\mathbf{y}_e+\mathbf{z}_e\right)}{2\pi}\right|\right|^2\right) \\
&= \prod_{e\in E,\mathfrak{X}_e}\left(1 - 8p(1-p)\left(\frac{\theta\left(\mathbf{y}_e+\mathbf{z}_e\right)}{2\pi}\right)^2\right) \text{ as } \mathbf{y}_e+\mathbf{z}_e = o\left(n^{1.26}\right),\ \theta = o\left(n^{-1.26}\right) \\
&\leqslant \prod_{e\in E,\mathfrak{X}_e}\left(1 - 8p(1-p)\left(\frac{\theta\left(p^3\cdot n^{1.245}\right)}{8\pi}\right)^2\right) \text{ as } \mathbf{Y}_e+\mathbf{Z}_e \geq \frac{p^3}{4}\cdot n^{1.245} \\
&\leqslant \left(1 - 8p(1-p)\left(\frac{\theta\left(p^3\cdot n^{1.245}\right)}{8\pi}\right)^2\right)^{\frac{8n}{10}} \\
&\leqslant e^{-\omega\left(\theta^2 n^{3.49}\right)}.
\end{aligned}$$

Combining this with Lemma 35 we have

$$\begin{aligned}
\mathbb{E}\left[e^{i\theta \mathbf{S}_n}\right] &= \mathbb{E}\left[\left|\mathbb{E}_E\left[e^{i\theta\sum_{e\in E}(\mathbf{Y}_e+\mathbf{Z}_e)\cdot \mathbf{I}_e}\right]\right|\right] \\
&= \left(\mathbb{P}(\mathfrak{X}^{\mathrm{c}}) + \max_{\mathbf{y}_e+\mathbf{z}_e|\mathfrak{X}\text{ occurs}} \mathbb{E}\left[\left|\mathbb{E}_E\left[e^{i\theta\sum_{e\in E}(\mathbf{Y}_e+\mathbf{Z}_e)\cdot \mathbf{I}_e}\right]\right|\right]\right)^{1/2} \\
&\leqslant e^{-\Theta(n^{0.4})} + e^{-\omega\left(\theta^2 n^{3.49}\right)},
\end{aligned}$$

which is at most $e^{-\Theta(n^{0.29})}$ for $\omega(n^{-1.6}) = \theta < n^{-1.4}$. We have thus shown that for all $\theta \in (n^{-1.96}, n^{1.4})$, the expectation

$$\mathbb{E}\left[\left|\mathbb{E}_E\left[e^{i\theta\sum_{e\in E}(\mathbf{Y}_e+\mathbf{Z}_e)\cdot \mathbf{I}_e}\right]\right|\right]$$

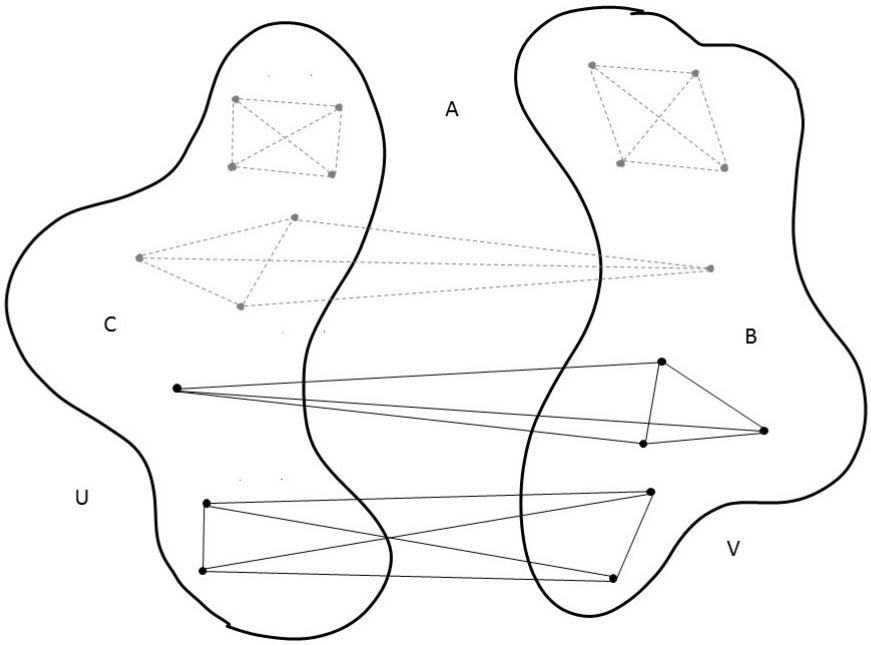
is bounded above by $e^{-\Theta(n^{0.06})}$ as required. ■

This proves Lemma 11.

2.7.5 Region $R_4 = \left(n^{-2.1}, n^{-1.96}\right)$

We now prove Lemma 10. As in the last section, we partition the vertex set $[n]$ into two parts, U and V. This time, we choose $|U| = |V| = \frac{n}{2}$ if n is even and $|U| = \frac{n+1}{2}$ and $|V| = \frac{n-1}{2}$ if n is odd. Let C denote the set of edges between vertices in U, let B denote

the set of edges between vertices in V and let A be the set of edges between a vertex in V and a vertex in U.



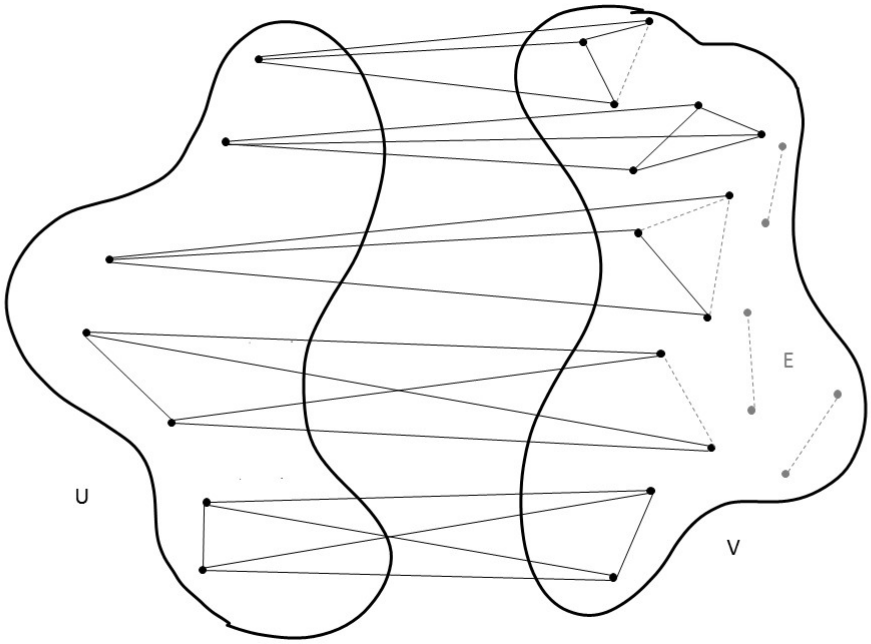
For the first part of this section, the setup is the same as in Section 2.7.4. By the same argument as in Section 2.7.4, so applying Lemma 18 to $\mathbf{S}_n$, we see that showing that the expectation $\mathbb{E}[\mathbf{S}_n]$ is small in absolute value again reduces to showing that for most choices of C, A and A' the expectation

$$\mathbf{\Psi} = \mathbb{E}_B\left[e^{i\theta\mathbf{\Phi}}\right]$$

is small in absolute value, where, as before

$$\mathbf{\Phi} = \sum_{u\in U}\left\langle\left(\mathbf{I}_u^{(3)}(A) - \mathbf{I}_u^{(3)}(A')\right), \mathbf{M}^{(3)}(B)\right\rangle + \sum_{\{u,w\}\in E(C)}\left\langle\left(\mathbf{I}_{u,w}^{(2)}(A) - \mathbf{I}_{u,w}^{(2)}(A')\right), \mathbf{M}^{(2)}(B)\right\rangle \ . \tag{2.70}$$

Now the setup deviates slightly from the last section. This time, we choose E to be a larger set of edges.



In particular, let E be a subset of edges in V, consisting of the set of all $\frac{k|V|}{2}$ edges in $\left\lfloor \frac{|V|}{2k} \right\rfloor$ disjoint copies of $K_{k,k}$, and denote by F the set of all edges inside V that are not contained in E. We now define $\mathbf{W}(A, A', C, F)$ to be the function equal to all terms in equation (2.70) that do not depend on E, so the function $\mathbf{W}(A, A', C, F)$ satisfies, where $(.)_e$ denotes the entry corresponding to the edge e in a vector and $\mathbf{I}_e$ is the indicator random variable of the presence of the edge e in G.

$$\begin{aligned}\mathbf{\Phi} = \mathbf{W}(A, A', C, F) &+ \sum_{e \in E} \sum_{u \in U} \sum_{v \in V} \left(\mathbf{I}_u^{(3)}(A) - \mathbf{I}_u^{(3)}(A')\right)_{e,v} \cdot \left(\mathbf{I}_v^{(2)}(F)\right)_e \cdot \mathbf{I}_e \\ &+ \sum_{e \in E} \sum_{\{u,w\} \in E(C)} \left(\mathbf{I}_{u,w}^{(2)}(A) - \mathbf{I}_{u,w}^{(2)}(A')\right)_e \cdot \mathbf{I}_e \\ &+ \frac{1}{2} \cdot \sum_{e=\{a,b\} \in E} \sum_{f=\{ac\} \in E} \sum_{u \in U} \left(\mathbf{I}_u^{(3)}(A) - \mathbf{I}_u^{(3)}(A')\right)_{abc} \cdot \left(\mathbf{I}_b^{(1)}(F)\right)_c \cdot \mathbf{I}_e \cdot \mathbf{I}_f \ , \end{aligned} \tag{2.71}$$

Note that we can split the sum $\mathbf{\Phi}$ up in the same way as in equation (2.53) in the last section, but this time there is an extra term accounting for the K_4 with multiple edges in E. As before, let

$$\begin{aligned}\mathbf{Y}_e &= \sum_{u \in U} \sum_{v \in V} \left(\mathbf{I}_u^{(3)}(A) - \mathbf{I}_u^{(3)}(A')\right)_{e,v} \cdot \left(\mathbf{I}_v^{(2)}(F)\right)_e \\ \mathbf{Z}_e &= \sum_{\{u,w\} \in E(C)} \left(\mathbf{I}_{u,w}^{(2)}(A) - \mathbf{I}_{u,w}^{(2)}(A')\right)_e \end{aligned}$$

and define

$$\mathbf{W}_{e,f} = \sum_{u \in U} \left(\mathbf{I}_u^{(3)}(A) - \mathbf{I}_u^{(3)}(A')\right)_{abc} \cdot \left(\mathbf{I}_b^{(1)}(F)\right)_c \ , \tag{2.72}$$

where $e = \{a, b\}$ and $f = \{a, c\}$ for some vertices $a, b, c \in V$ and note that none of $\mathbf{Y}_e$, $\mathbf{Z}_e$ and $\mathbf{W}_{e,f}$ depend on any of the edges in E. We write $e \sim f$ if two edges e and f share a vertex. Equation (2.71) can thus be rewritten as

$$\mathbf{\Phi} = \mathbf{W}(A, A', C, F) + \sum_{e \in E} (\mathbf{Y}_e + \mathbf{Z}_e) \cdot \mathbf{I}_e + \frac{1}{2} \cdot \sum_{e \in E} \sum_{f \sim e} \mathbf{W}_{e,f} \cdot \mathbf{I}_e \cdot \mathbf{I}_f \ . \tag{2.73}$$

Note that the only terms depending on edges inside E are the random variables $\mathbf{I}_e$ and so only the final terms in the sum depend on more than one of the edges inside E. Then

$$\begin{aligned}
|\mathbf{\Psi}| &= \left| \mathbb{E}_{E,F} \left[e^{i\theta \mathbf{\Phi}} \right] \right| \\
&\leqslant \mathbb{E}_F \left[\left| \mathbb{E}_E \left[e^{i\theta \mathbf{\Phi}} \right] \right| \right] \\
&= \mathbb{E}_F \left[\left| \mathbb{E}_E \left[e^{i\theta \left(\mathbf{W}(A,A',C,F) + \sum_{e \in E} (\mathbf{Y}_e + \mathbf{Z}_e) \cdot \mathbf{I}_e + \frac{1}{2} \sum_{e \in E} \sum_{f \sim e} \mathbf{W}_{e,f} \cdot \mathbf{I}_e \cdot \mathbf{I}_f \right)} \right] \right| \right] \\
&= \mathbb{E}_F \left[\left| e^{i\theta \cdot \mathbf{W}(A,A',C,F)} \cdot \mathbb{E}_E \left[e^{i\theta \left(\sum_{e \in E} (\mathbf{Y}_e + \mathbf{Z}_e) \cdot \mathbf{I}_e + \frac{1}{2} \sum_{e \in E} \sum_{f \sim e} \mathbf{W}_{e,f} \cdot \mathbf{I}_e \cdot \mathbf{I}_f \right)} \right] \right| \right] \\
&= \mathbb{E}_F \left[\left| \mathbb{E}_E \left[e^{i\theta \left(\sum_{e \in E} (\mathbf{Y}_e + \mathbf{Z}_e) \cdot \mathbf{I}_e + \frac{1}{2} \sum_{e \in E} \sum_{f \sim e} \mathbf{W}_{e,f} \cdot \mathbf{I}_e \cdot \mathbf{I}_f \right)} \right] \right| \right]
\end{aligned}$$

and so the problem of bounding the absolute value of $\mathbf{\Psi}$ reduces to showing that the expectation $\mathbf{\Xi} = \mathbb{E}_E \left[e^{i\theta \left(\sum_{e \in E} (\mathbf{Y}_e + \mathbf{Z}_e) \cdot \mathbf{I}_e + \frac{1}{2} \cdot \sum_{e \in E} \sum_{f \sim e} \mathbf{W}_{e,f} \cdot \mathbf{I}_e \cdot \mathbf{I}_f \right)} \right]$ is small in absolute value for most choices of F. We will later condition on all edges outside E. After conditioning on these, $\mathbf{\Phi}$ is just a constant plus the sum of $|E|$ Bernoulli random variables each multiplied by a constant plus the sum of some cross terms of these each multiplied by other constants. So as long as we can show that the influence of the cross terms is not too large and the constants we multiply the Bernoulli random variables by are roughly of size $n^{1.5}$ for the range of θ we are aiming for, we can bound the expectation of the characteristic function as outlined in Section 2.6. In the following we will denote an observation of $\mathbf{Y}_e$ by $\mathbf{y}_e$ and an observation of $\mathbf{Z}_e$ by $\mathbf{z}_e$.

Let $r = \frac{|V|}{2k}$ and for each $i \in \{1, \ldots, r\}$, let E_i be the set of edges in E formed by the i^{th} copy of $K_{k,k}$. If we then define $\mathbf{K}_j = \mathbb{E}_{E_j} \left[\sum_{e \in E_j} \sum_{f \sim e} \mathbf{W}_{e,f} \cdot \mathbf{I}_e \cdot \mathbf{I}_f \right]$ and apply Lemma 16 we get that, taking all edges outside E as fixed

$$\begin{aligned}
|\mathbf{\Xi}| &= \left| \mathbb{E}_{E_1, \ldots, E_r} \left[e^{i\theta \left(\sum_{e \in E_j} (\mathbf{Y}_e + \mathbf{Z}_e) \cdot \mathbf{I}_e + \frac{1}{2} \sum_{e \in E_j} \sum_{f \sim e} \mathbf{W}_{e,f} \cdot \mathbf{I}_e \cdot \mathbf{I}_f \right)} \right] \right| \\
&= \left| \prod_{j=1}^{r} \mathbb{E}_{E_j} \left[e^{i\theta \left(\sum_{e \in E_j} (\mathbf{Y}_e + \mathbf{Z}_e) \cdot \mathbf{I}_e + \frac{1}{2} \sum_{e \in E_j} \sum_{f \sim e} \mathbf{W}_{e,f} \cdot \mathbf{I}_e \cdot \mathbf{I}_f \right)} \right] \right| \\
&\leqslant \prod_{j=1}^{r} \left(\left| \mathbb{E}_{E_j} \left[e^{i\theta \left(\sum_{e \in E} (\mathbf{Y}_e + \mathbf{Z}_e) \cdot \mathbf{I}_e \right)} \right] \right| + \mathbb{E}_{E_j} \left[\left| e^{\frac{i\theta}{2} \left(\sum_{e \in E_j} \sum_{f \sim e} \mathbf{W}_{e,f} \cdot \mathbf{I}_e \cdot \mathbf{I}_f - \mathbf{K}_j \right)} - 1 \right| \right] \right) .
\end{aligned} \tag{2.74}$$

For each E_j for $j = 1, \dots, r$ we now bound the two expectations separately. To bound $\left|\mathbb{E}_{E_j}\left[e^{i\theta\left(\sum_{e\in E}(\mathbf{Y}_e+\mathbf{Z}_e)\cdot\mathbf{I}_e\right)}\right]\right|$ we use essentially the same argument as in the previous section. Note that all definitions are exactly the same as in the previous section, only the sets E and F have changed and as a result the events $\mathbf{Y}_e$ and $\mathbf{Z}_e$. As in the last section, for an edge $e \in E$ let $\mathscr{Y}_e$ be the event that $\mathbf{Y}_e$ is of roughly the right size, so in particular the event that

$$\mathbf{Y}_e \in \left[\frac{p^3}{4}|U|^{0.49}|V|, \frac{9p^3}{4}|U|^{0.51}|V|\right] .$$

Similarly, we define the event $\mathscr{Z}_e$ to be the event that $\mathbf{Z}_e$ is roughly of the right size, so in particular the event that

$$\mathbf{Z}_e \in \left[\frac{p^3}{8}|U|^{1.49}, \frac{9p^3}{4}|U|^{1.51}\right] .$$

The event $\boldsymbol{x}_e$, which we will show to hold with high probability for at least $\frac{4}{10}$ of all edges $e \in E_j$ for at least $\frac{1}{2}$ of all $j \in \{1, \dots, r\}$ is the event that the difference between the number of common neighbours of the endvertices of e between A and A' is roughly of size $\sqrt{|U|}$, which is how large we expect it to be on average. Formally, we define $\boldsymbol{x}_e$ to be the event that

$$\boldsymbol{x}_e := \left\{\sum_{u\in U}(\mathbf{I}_u^{(2)}(A) - \mathbf{I}_u^{(2)}(A'))_e \in [|U|^{0.49}, |U|^{0.51}]\right\} .$$

Let

$$T = \{e \in E \mid \boldsymbol{x}_e \text{ occurs}\} , \tag{2.75}$$

define $\boldsymbol{x}_{E_j}$ to be the event that $|T \cap E_j| \geq \frac{4}{10}|E_j|$, and let $\boldsymbol{x}$ be the event that $\boldsymbol{x}_{E_j}$ holds for at least $\frac{1}{2}$ of all $j \in \{1, \cdots, r\}$. For an edge $e \in E$, we let $\mathscr{X}_e$ be the event $\mathscr{Y}_e \cap \mathscr{Z}_e$. The event $\mathscr{X}$ is then defined to be the event that $\mathscr{X}_e$ holds for at least $\frac{4}{10}$ of all edges $e \in E_j$ for at least $\frac{1}{2}$ of all $j \in \{1, \cdots, r\}$. We wish to show that the event $\mathscr{X}$ holds with high probability.

The only two events left to be defined are the events $\mathscr{y}_e$ and $\mathscr{z}_e$. For a fixed edge $e \in E$, let V_e denote the set of common neighbours in F of the endvertices of e that lie in V, and let m_e denote the size of V_e. We define $\mathscr{y}_e$ to be the event that m_e is of roughly the size we expect it to be, so $\mathscr{y}_e$ is the event that

$$m_e \in \left[\frac{p^2|V|}{2}, \frac{3p^2|V|}{2}\right]$$

and let $\mathscr{y}$ be the event that $\mathscr{y}_e$ holds for all $e \in E$.

For an edge $e \in E$, let S_e denote the set of vertices in U that are connected to both endvertices of e in A. Similarly, let S'_e denote the set of vertices in U that are connected

to both endvertices of e in A'. For a set M of vertices, we further denote by $\mathscr{C}(M)$ be the number of connected pairs of vertices in M. For an edge $e \in E$, we let $\bar{\varkappa}_e$ be the event that the number of common neighbours of the endvertices of e in U is of roughly the size we expect it to be, more precisely, $\bar{\varkappa}_e$ is the event that

$$|S_e| \in \left[\frac{p^2}{2}|U|, \frac{3p^2}{2}|U|\right] \ . \tag{2.76}$$

Furthermore, $\bar{\varkappa}$ is the event that $\bar{\varkappa}_e$ holds for all $e \in E$.

Having defined these events, we now go on to bound the probability of $\mathfrak{X}$ occurring. We start by bounding the probability of the event $\bar{\boldsymbol{x}}$.

Lemma 37. *We have that* $\mathbb{P}(\bar{\boldsymbol{x}}) \geq 1 - e^{-\Theta(r)}$.

Proof. Fix a $j \in \{1, \dots, r\}$. The proof of Lemma 30 in the last section assumes independence of the common neighbours between edges in E. This does not hold in the setup of this section since some of the edges inside E_j share vertices. As in the last section however, for a fixed edge e, the sum $\sum_{u \in U}(\mathbf{I}_u^{(2)}(A) - \mathbf{I}_u^{(2)}(A'))_e$ is the sum of $|U|$ random variables, each of which is equal to 1 with probability $p^2(1-p^2)$, 0 with probability $1 - 2(p^2 - p^4)$ and equal to -1 with probability $p^2(1-p^2)$ independently of all others. By standard concentration and anti-concentration estimates for binomial distributions, we have that $\mathbb{P}(\bar{\boldsymbol{x}}_e) \geq 0.45$, as long as n is sufficiently large, depending on p. Further, let $\mathbf{X}_e$ denote the indicator random variable for the occurrence of the event $\bar{\boldsymbol{x}}_e$.

Now note that within E_j, the random variable $\mathbf{X}_e$ depends on exactly $2k$ other events $\mathbf{X}_f$ for some edges f and is independent of the rest. We denote by $\mathbf{X}_{E_j}$ the number of edges $e \in E_j$ for which $\bar{\boldsymbol{x}}_e$ holds, so $\mathbf{X}_{E_j} = \sum_{e \in E_j} \mathbf{X}_e$. Then $\mathbb{E}[\mathbf{X}_{E_j}] = k^2\mathbb{P}(\bar{\boldsymbol{x}}_e)$ and

$$\begin{aligned}
\operatorname{Var}(\mathbf{X}_{E_j}) &= \sum_{e \in E_j} \operatorname{Var}(\mathbf{X}_e) + \sum_{e \neq f} \operatorname{Cov}(\mathbf{X}_e, \mathbf{X}_f) \\
&\leqslant |E_j| + |E_j|k \\
&\leqslant 3k^3 \ .
\end{aligned}$$

We now apply Chebyshev's inequality to $\mathbf{X}_{E_j}$ to obtain the following bound on the probability of $\bar{\boldsymbol{x}}_{E_j}^{\mathrm{c}}$ occurring

$$\begin{aligned}
\mathbb{P}(\bar{\boldsymbol{x}}_{E_j}^{\mathrm{c}}) &= \mathbb{P}\left(\mathbf{X}_{E_j} < \frac{4}{10}k^2\right) \\
&= \mathbb{P}\left(\mathbb{E}[\mathbf{X}_{E_j}] - \mathbf{X}_{E_j} > \frac{k^2}{20}\right) \\
&= O\left(\frac{1}{k}\right) \ .
\end{aligned}$$

Since the different blocks E_j do not share any vertices, the events $\boldsymbol{x}_{E_j}$ are independent for $j \in \{1, \dots, r\}$. The number of blocks E_j for which $\boldsymbol{x}_{E_j}$ holds is hence Binomially distributed with parameters $r = \frac{|V|}{2k}$ and $\mathbb{P}(\boldsymbol{x}_{E_j}) = 1 - \Theta\left(\frac{1}{k}\right)$. For k large enough and by using the Chernoff bound from Theorem 21, the event $\boldsymbol{x}_{E_j}$ thus holds for at least 4/10 of all blocks in E with probability $1 - e^{-\Theta(r)}$. Thus the good event η holds with probability at least $1 - e^{-\Theta(r)}$, as required. ■

Lemma 38. *We have that* $\mathbb{P}(\boldsymbol{y}) \geq 1 - 2r|V|e^{-\Theta(|V|)}$.

Proof. The proof is similar to the proof of Lemma 31 in the previous section. For a fixed edge $e \in E$, the vertex $v \in V$ is a common neighbour in F of the endvertices of e with probability p^2. However, note that the vertex v must not be contained in the same $K_{k,k}$ as e, since otherwise one of the edges connecting e to v is contained in E. As a result, there are at most $|V| - 2k$ possible common neighbours connected to e only by edges in F. The appearances of $v \in V$ and $w \in V$ as common neighbours of e are independent as they depend on disjoint sets of edges. The random variable V_e is hence distributed as a Binomial random variable with parameters $|V| - 2k$ and p^2 plus at most $2k$ common neighbours connected by edges in E, which we do not want to count. By the Chernoff Bound from Theorem 21 and using that $k = o(|V|)$, the edge e has at least $\frac{p^2|V|}{2}$ and at most $\frac{3p^2|V|}{2}$ common F-neighbours in V with probability at least $1 - e^{-\Theta(|V|)}$. By a union bound applied to the complements, we have that $\boldsymbol{y}_e$ holds for all $e \in E$ with probability at least $1 - |E|e^{-\Theta(|V|)}$. Thus $\mathbb{P}(\boldsymbol{y}) \geq 1 - 2r|V|e^{-\Theta(|V|)}$, as required. ■

Lemma 39. *We have that* $\mathbb{P}(\boldsymbol{z}) \geq 1 - |E|e^{-\Theta(|U|)}$.

Proof. The same proof as in Lemma 32 applies. ■

From now on, we condition on the events $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}$ occurring. Conditional on these, we wish to show that for all $e \in T$, the random variables $\mathbf{Y}_e$ and $\mathbf{Z}_e$ are with high probability of roughly the size we expect them to be.

Lemma 40. *For a fixed edge $e \in E$, we have that* $\mathbb{P}\left(\boldsymbol{\mathcal{Y}}_e^{\mathrm{c}} \mid \boldsymbol{x}_e, \boldsymbol{y}_e\right) \leqslant e^{-\Theta(n^{0.48})}$.

Proof. The same proof as in Lemma 33 applies. ■

So by applying a union bound on the complements, we have that conditional on $\boldsymbol{x}$ and $\boldsymbol{y}$ with probability at least $1 - |V|e^{-\Theta(n^{0.48})}$ the event $\boldsymbol{\mathcal{Y}}_e$ holds for all $e \in T$. We now need to show that for all $e \in T$, $\mathbf{Z}_e$ is also of roughly the size we expect it to be. Recall that $\mathbf{Z}_e$ is the event that

$$\mathbf{Z}_e \in \left[\frac{p^2}{8}|U|^{1.49}, \frac{9p^3}{4}|U|^{1.51}\right] .$$

We now go on to bound the probability of $\boldsymbol{\mathcal{Z}}_e$ occurring conditional on the events $\boldsymbol{x}_e$ and $\boldsymbol{z}_e$.

Lemma 41. *For any fixed edge $e \in E$ we have that $\mathbb{P}(\mathfrak{X}_e^{\mathrm{c}} \mid \boldsymbol{x}_e, \boldsymbol{z}_e) \leqslant e^{-\Theta\left(|U|^{0.98}\right)}$.*

Proof. The same proof as in Lemma 34 holds. ■

So by applying a union bound to the complements of $\mathfrak{X}_e$ over all $e \in T$ and also conditioning on $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}$ occurring, we have that with probability at least $1-k|V|e^{-\Theta\left(|U|^{0.98}\right)}$ the event $\mathfrak{X}$ occurs. By another union bound applied to the complements, we see that with probability at least

$$1 - k|V|e^{-\Theta\left(|U|^{0.98}\right)} - e^{-\Theta(n^{0.48})}$$

the events $\mathcal{Y}_e$ and $\mathfrak{X}_e$ both occur for all $e \in T$ and further $|E_j \cap T| \geq \frac{4}{10}|E_j|$ for at least $\frac{1}{2}$ of all $j \in \{1, \cdots, r\}$. Recall that this is equivalent to the event $\mathfrak{X}$ occurring. We can finally bound the probability of $\mathfrak{X}$ occurring.

Lemma 42. *We have that $\mathbb{P}(\mathfrak{X}) \geq 1 - e^{-\Theta(n^{0.4})}$.*

Proof. The proof for this is essentially the same as the proof of Lemma 35. Conditional on the events $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}$ occurring, we have that the event $\mathfrak{X}$ occurs with probability at least $1 - k|V|e^{-\Theta\left(|U|^{0.98}\right)} - e^{-\Theta(n^{0.48})}$. Thus

$$\begin{aligned}
\mathbb{P}\left(\mathfrak{X}^{\mathrm{c}}\right) &= \mathbb{P}\left(\mathfrak{X}^{\mathrm{c}} \cap \boldsymbol{x} \cap \boldsymbol{y} \cap \boldsymbol{z}\right) + \mathbb{P}\left(\mathfrak{X}^{\mathrm{c}} \cap \left(\boldsymbol{x}^{\mathrm{c}} \cup \boldsymbol{y}^{\mathrm{c}} \cup \boldsymbol{z}^{\mathrm{c}}\right)\right) \\
&\leqslant \mathbb{P}\left(\mathfrak{X}^{\mathrm{c}} \cap \boldsymbol{x} \cap \boldsymbol{y} \cap \boldsymbol{z}\right) + \mathbb{P}\left(\boldsymbol{x}^{\mathrm{c}} \cup \boldsymbol{y}^{\mathrm{c}} \cup \boldsymbol{z}^{\mathrm{c}}\right) \\
&\leqslant \left(|V|e^{-\Theta\left(|U|^{0.98}\right)} + e^{-\Theta(n^{0.48})}\right) + \left(e^{-\Theta(r)} + 2r|V|e^{-\Theta(|V|)} + |T|e^{-\Theta(|U|)}\right) \\
&\leqslant e^{-\Theta(n^{0.4})} + e^{-\Theta(r)} \,,
\end{aligned}$$

as we have $|U|, |V| = \Theta(n)$. Equivalently then

$$\mathbb{P}\left(\mathfrak{X}\right) \geq 1 - e^{-\Theta(n^{0.4})} \,,$$

as required. ■

We have shown that both $\mathcal{Y}_e$ and $\mathfrak{X}_e$ are with high probability contained in a small interval. Now define all blocks j such that $|E_j \cap T| \geq \frac{4}{10}|E_j|$ to be 'good', so these are the blocks where both $\mathcal{Y}_e$ and $\mathfrak{X}_e$ hold for a large fraction of the vertices. We use the fact that both $\mathcal{Y}_e$ and $\mathfrak{X}_e$ hold to find a bound on

$$\left|\mathbb{E}_{E_j}\left[e^{i\theta\sum_{e\in E_j}(\mathbf{y}_e+\mathbf{z}_e)\cdot\mathbf{I}_e}\right]\right| \tag{2.77}$$

for all 'good' j.

Lemma 43. *For all 'good' j and for all $\theta \in (n^{-2.1}, n^{-1.96})$ we have*

$$\left|\mathbb{E}_{E_j}\left[e^{i\theta\mathbf{S}_{E_j}}\right]\right| \leqslant e^{-\Theta(\theta^2 n^{2.98}k^2)} \,, \tag{2.78}$$

where $\mathbf{S}_{E_j} := \sum_{e\in E_j}(\mathbf{y}_e + \mathbf{z}_e)\cdot\mathbf{I}_e$.

Proof. Recall that $|U| = |V| = \frac{n}{2}$ if n is even and $|U| = \frac{n-1}{2}$, $|V| = \frac{n+1}{2}$ if n is odd. Then $\mathbf{S}_{E_j}$ is the sum of $|E_j|$ independent Bernoulli random variables each multiplied by a constant. Since we have assumed that the event $\mathfrak{X}_{E_j}$ occurs, at least 4/10 of these constants lie in the interval $[\frac{p^3}{8}n^{1.49}, \frac{5}{2}n^{1.51}]$ for E_j. This allows us to bound the conditional expectation as follows when the event $\boldsymbol{x}_{E_j}$ occurs

$$\begin{aligned}
\left|\mathbb{E}_{E_j}\left[e^{i\theta\mathbf{S}_{E_j}}\right]\right| &= \left|\mathbb{E}_{E_j}\left[\prod_{e\in E_j} e^{i\theta(\mathbf{y}_e+\mathbf{z}_e)\cdot\mathbf{I}_e}\right]\right| \\
&\leqslant \prod_{e\in E_j}\left|\mathbb{E}_{E_j}\left[e^{i\theta(\mathbf{y}_e+\mathbf{z}_e)\cdot\mathbf{I}_e}\right]\right| \text{ since } \mathbf{I}_e \text{ are independent} \\
&\leqslant \prod_{e\in E_j,\mathfrak{X}_e}\left(1-8p(1-p)\left|\left|\frac{\theta\left(\mathbf{y}_e+\mathbf{z}_e\right)}{2\pi}\right|\right|^2\right) \\
&= \prod_{e\in E_j,\mathfrak{X}_e}\left(1-8p(1-p)\left(\frac{\theta\left|\mathbf{y}_e+\mathbf{z}_e\right|}{2\pi}\right)^2\right) \text{ as } |\mathbf{y}_e+\mathbf{z}_e| \leqslant \frac{5n^{1.51}}{2}, \theta = o\left(n^{-1.75}\right) \\
&\leqslant \prod_{e\in E_j,\mathfrak{X}_e}\left(1-8p(1-p)\left(\frac{\theta\left(\frac{p^3}{8}n^{1.49}\right)}{2\pi}\right)^2\right) \text{ as } |\mathbf{y}_e+\mathbf{z}_e| \geqslant \frac{p^3n^{1.49}}{8} \\
&\leqslant \left(1-8p(1-p)\left(\frac{\theta\left(p^3n^{1.49}\right)}{16\pi}\right)^2\right)^{|T\cap E_j|} \\
&= e^{-\Theta(\theta^2n^{2.98}\,|T\cap E_j|)} \\
&= e^{-\Theta(\theta^2n^{2.98}k^2)} \text{ as we assume that the event } \boldsymbol{x}_{E_j} \text{ holds,}
\end{aligned} \tag{2.79}$$

as required. ■

We are thus left with bounding the second expectation in (2.74), namely

$$\mathbb{E}_{E_j}\left[\left|e^{i\theta\left(\sum_{e\in E_j}\sum_{f\sim e}\mathbf{W}_{e,f}\cdot\mathbf{I}_e\cdot\mathbf{I}_f-\mathbf{K}_j\right)}-1\right|\right] .$$

We do so by showing that the sum $\sum_{e\in E_j}\sum_{f\sim e}\mathbf{W}_{e,f}\cdot\mathbf{I}_e\cdot\mathbf{I}_f$ and its expectation $\mathbf{K}_j = \mathbb{E}_{E_j}\left[\sum_{e\in E_j}\sum_{f\sim e}\mathbf{W}_{e,f}\cdot\mathbf{I}_e\cdot\mathbf{I}_f\right]$ are small with high probability, so they lie close together with high probability. To be able to bound these expectations, we need to condition on one further event. Let $\mathscr{W}_{E_j}$ be the event that for E_j

$$\sum_{e,f\in E_j}|\mathbf{W}_{e,f}| \leqslant k^3\cdot|U|^{0.51} \tag{2.80}$$

and thus when $\mathscr{W}_{E_j}$ holds

$$\mathbb{E}\left[\sum_{e,f\in E_j}\mathbf{W}_{e,f}\cdot\mathbf{I}_e\cdot\mathbf{I}_f\right] \leqslant k^3\cdot|U|^{0.51} . \tag{2.81}$$

The event we will condition on, $\mathscr{W}$, is then the event that $\mathscr{W}_{E_j}$ occurs for all $j \in \{1, \cdots, r\}$. We now bound the probability of $\mathscr{W}$ occurring.

Lemma 44. *We have that* $\mathbb{P}(\mathscr{W}) \geqslant 1 - rk^3 e^{-\Theta(|U|^{0.02})}$.

Proof. Firstly note that by Theorem 22

$$\mathbb{P}(\mathbf{W}_{e,f} \leqslant |U|^{0.51}) \geqslant 1 - e^{\Theta(|U|^{0.02})} \ . \tag{2.82}$$

Furthermore, if the edges e and f do not share a vertex, then

$$W_{e,f} = 0 \ . \tag{2.83}$$

In a single copy of E_j, there are at most k^3 pairs of edges e, f such that e and f share a vertex. Since further $\mathbf{I}_e$ and $\mathbf{I}_f$ only take values in $\{0, 1\}$ we have that for any j,

$$\mathbb{P}(W_{E_j}) \geqslant 1 - k^3 e^{-\Theta(|U|^{0.02})} \ . \tag{2.84}$$

Thus by applying a union bound over the complements we get

$$\mathbb{P}(\mathscr{W}) \geqslant 1 - rk^3 e^{-\Theta(|U|^{0.02})} \ , \tag{2.85}$$

as required. ■

We now show that conditional on $\mathscr{W}$, the second set of expectations are also small in value.

Lemma 45. *When* $\mathscr{W}$ *occurs,*

$$\mathbb{E}_{E_j}\left[\left|e^{i\theta\left(\sum_{e\in E_j}\sum_{f\sim e}\mathbf{w}_{e,f}\cdot\mathbf{I}_e\cdot\mathbf{I}_f - \mathbf{K}_j\right)} - 1\right|\right] \leqslant 2\theta k^3 n^{0.51} \ , \tag{2.86}$$

for all j.

Proof. By using the fact that for any real number α we have that $|e^{i\alpha} - 1| \leqslant 2\pi \left|\frac{\alpha}{2\pi}\right|$, we can bound the expectation as follows:

$$\begin{aligned}
\mathbb{E}_{E_j}\left[\left|e^{i\theta\left(\sum_{e\in E_j}\sum_{f\sim e}\mathbf{w}_{e,f}\cdot\mathbf{X}_e\cdot\mathbf{X}_f - \mathbf{K}_j\right)} - 1\right|\right] &\leqslant 2\pi\, \mathbb{E}_{E_j}\left[\left|\frac{\theta\left(\sum_{e\in E_j}\sum_{f\sim e}\mathbf{w}_{e,f}\cdot\mathbf{I}_e\cdot\mathbf{I}_f - \mathbf{K}_j\right)}{2\pi}\right|\right] \\
&\leqslant 4\theta k^3 |U|^{0.51} \text{ since } \mathscr{W} \text{ occurs} \\
&= \Theta\left(\theta k^3 n^{0.51}\right) \ ,
\end{aligned} \tag{2.87}$$

as required. ■

Substituting the bounds from equations (2.79) and (2.87) into equation (2.74) and assuming that both $\mathfrak{X}$ and $\mathscr{W}$ hold, we see that, conditional on all edges outside E we have

$$\begin{aligned}
|\boldsymbol{\Xi}| &= \prod_{j=1}^{r} \left| \mathbb{E}_{E_j} \left[e^{i\theta \left(\sum_{e \in E_j} (\mathbf{Y}_e + \mathbf{Z}_e) \cdot \mathbf{I}_e + \frac{1}{2} \sum_{e \in E_j} \sum_{f \sim e} \mathbf{W}_{e,f} \cdot \mathbf{I}_e \cdot \mathbf{I}_f \right)} \right] \right| \\
&\leqslant \prod_{j \in \{1,\ldots,l\}, x_{E_j}} \left(\left| \mathbb{E}_{E_j} \left[e^{i\theta \left(\sum_{e \in E} (\mathbf{y}_e + \mathbf{z}_e) \cdot \mathbf{X}_e \right)} \right] \right| + \mathbb{E}_{E_j} \left[\left| e^{i\theta \left(\sum \sum \mathbf{w}_{e,f} \cdot \mathbf{X}_e \cdot \mathbf{X}_f - \mathbf{K}_j \right)} - 1 \right| \right] \right) \\
&\leqslant \prod_{j \in \{1,\ldots,l\}, x_{E_j}} \left(e^{-\Theta(\theta^2 n^{2.98} \cdot k^2)} + 2\theta k^3 n^{0.51} \right) \\
&\leqslant \left(e^{-\Theta(\theta^2 n^{2.98} \cdot k^2)} + 2\theta k^3 n^{0.51} \right)^{\Theta(r)} .
\end{aligned}$$

Thus taking the expectation of $|\boldsymbol{\Xi}|$ we see that

$$\begin{aligned}
\mathbb{E}[|\boldsymbol{\Xi}|] &\leqslant \mathbb{E}[|\boldsymbol{\Xi}| \mid \mathfrak{X} \cap \mathscr{W}] \cdot \mathbb{P}(\mathfrak{X} \cap \mathscr{W}) + \mathbb{E}[|\boldsymbol{\Xi}| \mid \mathfrak{X}^{\mathrm{c}} \cup \mathscr{W}^{\mathrm{c}}] \cdot \mathbb{P}(\mathfrak{X}^{\mathrm{c}} \cup \mathscr{W}^{\mathrm{c}}) \\
&\leqslant \left(e^{-\Theta(\theta^2 n^{2.98} \cdot k^2)} + 2\theta k^3 n^{0.51} \right)^{\Theta(r)} + e^{-\Theta(n^{0.04})} + e^{-\Theta(r)} + r k^3 e^{-\Theta(|U|^{0.02})} . \quad (2.88)
\end{aligned}$$

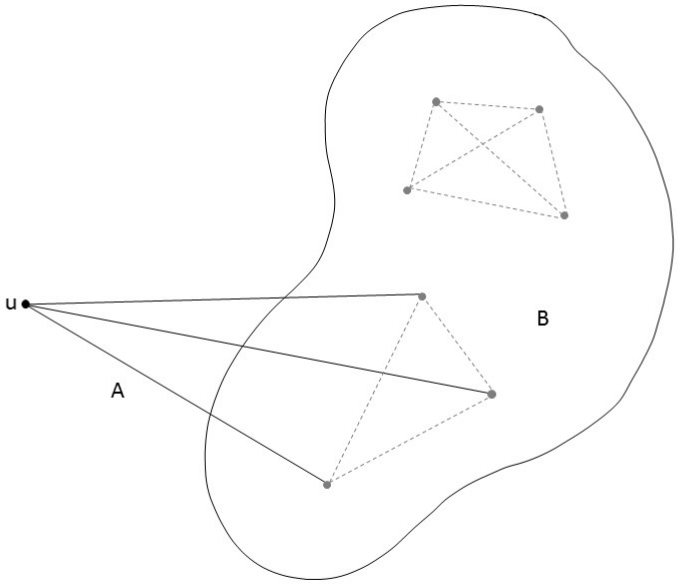
We can hence bound the expectation of the characteristic function. For simplicity of notation, we let $\mathbf{D}(A, A', C) = \left(\sum_{\{u,v,w\} \in \triangle(C)} \left\langle \left(\mathbf{I}^{(1)}_{u,v,w}(A) - \mathbf{I}^{(1)}_{u,v,w}(A') \right), \mathbf{I} \right\rangle \right)$ be a function of A, A' and C only. Then

$$\begin{aligned}
\left| \mathbb{E} \left[e^{i\theta \mathbf{S}_n} \right] \right|^2 &\leqslant \mathbb{E}_{A,A',C} \left[e^{i\theta \cdot \mathbf{D}(A,A',C)} \, \mathbb{E}_F \left[\left| \mathbb{E}_E \left[e^{i\theta \left(\sum_{e \in E} (\mathbf{Y}_e + \mathbf{Z}_e) \cdot \mathbf{X}_e + \sum_{e \in E_j} \sum_{f \sim e} \mathbf{W}_{e,f} \cdot \mathbf{X}_e \cdot \mathbf{X}_f \right)} \right] \right| \right] \right] \\
&\leqslant \mathbb{E}_{A,A',C} \left[\left| e^{i\theta \cdot \mathbf{D}(A,A',C)} \, \mathbb{E}_F \left[|\boldsymbol{\Xi}| \right] \right| \right] \\
&\leqslant \mathbb{E}_{A,A',C} \left[\mathbb{E}_F \left[\left(e^{-\Theta(\theta^2 n^{2.98} \cdot k^2)} + 2\theta k^3 n^{0.51} \right)^{\Theta(r)} + e^{-\Theta(n^{0.04})} + e^{-\Theta(r)} + r k^3 e^{-|U|^{0.02}} \right] \right] .
\end{aligned}$$

If we choose $k = n^{0.4}$, then $r = \frac{|V|}{2k} = \Theta(n^{0.6})$ and $\left| \mathbb{E} \left[e^{i\theta \mathbf{S}_n} \right] \right|$ is bounded above by $\Theta(e^{-n^{0.02}})$ for all θ in the range $n^{-2.1} \leqslant \theta \leqslant n^{-1.96}$. This proves Lemma 10.

2.7.6 Region $R_7 = \left(n^{-0.6}, \pi \right)$

We now prove Lemma 13. As before, we bound the characteristic function by using a conditioning argument. The setup in this section is as follows: we partition the vertex set $[n]$ into two sets, $U = \{u\}$ and $V = \binom{[n]}{2} \setminus U$. Further, denote the set of edges between u and V by A and the set of edges within V by B.



As in the previous sections, we denote by $\mathbf{S}_n$ the number of K_4 in G and by $\mathbf{I}_u^{(3)}(A)$ the vector of size $\binom{|V|}{3}$ with one entry for each subset of V of size three, where the entry corresponding to $\{i, j, k\}$ is equal to one if i, j and k are all connected to u in G and equal to zero otherwise. The vector $\mathbf{M}^{(3)}(B)$ is defined to be the vector of size $\binom{|V|}{3}$ with one entry for each subset of V of size three, where the entry corresponding to $\{i, j, k\}$ is equal to one if and only if all edges between $\{i, j, k\}$ exist and zero otherwise. $\mathbf{M}^{(4)}(B)$ is defined similarly for sets of vertices in V of size four and $\mathbf{I}$ is defined to be the vector with ones in each entry. Then the number of K_4 in G is equal to the number of K_4 with one vertex or zero vertices inside U, so

$$\mathbf{S}_n = \left\langle \mathbf{I}_u^{(3)}(A), \mathbf{M}^{(3)}(B) \right\rangle + \left\langle \mathbf{M}^{(4)}(B), \mathbf{I} \right\rangle \ . \tag{2.89}$$

By applying Lemma 17 to $\mathbf{S}_n$ we thus have

$$\left| \mathbb{E}\left[e^{i\theta \mathbf{S}_n} \right] \right|^2 \leq \mathbb{E}_{A,A'} \left[\mathbb{E}_B \left[e^{i\theta \left\langle \mathbf{I}_u^{(3)}(A) - \mathbf{I}_u^{(3)}(A'), \mathbf{M}^{(3)}(B) \right\rangle} \right] \right] \ . \tag{2.90}$$

Since we aim to show that the expectation $\mathbb{E}\left[e^{i\theta \mathbf{S}_n} \right]$ is small in absolute value, equation (2.90) implies that it is enough to show that for most choices of A and A', the expectation

$$\mathbf{X} = \mathbb{E}_B \left[e^{i\theta \Psi} \right] \ ,$$

where $\Psi = \left\langle \mathbf{I}_u^{(3)}(A) - \mathbf{I}_u^{(3)}(A'),\, \mathbf{M}^{(3)}(B) \right\rangle$, is small in absolute value. Note that the random vector $\left(\mathbf{I}_u^{(3)}(A) - \mathbf{I}_u^{(3)}(A') \right)$ has the following distribution: Suppose $J \subseteq [n] \setminus \{u\}$ is a random subset where each $v \in [n] \setminus \{u\}$ occurs with probability p and let J' be an

independent copy of J. Then the $\{i,j,k\}$ coordinate of $\left(\mathbf{I}_u^{(3)}(A) - \mathbf{I}_u^{(3)}(A')\right)$ is equal to 1 if $\{i,j,k\} \subseteq J$ and $\{i,j,k\} \not\subseteq J'$, equal to -1 if $\{i,j,k\} \subseteq J'$ and $\{i,j,k\} \not\subseteq J$ and equal to 0 otherwise. Let α be the event that the symmetric difference between J and J' has size at least $np(1-p)$ and let β be the event that there exists a vertex v in $[n] \setminus \{u\}$ such that u is connected to v in both A and A'. Call a pair (A, A') "good" if the events α and β occur simultaneously. We define this event to be the event η, so $\eta = \alpha \cap \beta$. We now bound the probability that the pair (A, A') is "good". For simplicity of notation, we define the set of vertices $v \in V$ such that uv lies in the symmetric difference between A and A' by $D(A, A')$ and the set of vertices $v \in V$ such that uv is present in A and not in A' by $D_A(A, A')$.

Lemma 46. *We have that* $\mathbb{P}(\alpha) \geq 1 - e^{-\Theta(n)}$.

Proof. Note that a fixed vertex $v \in V$ is contained in $D(A, A')$ with probability $2p(1-p)$ independently of all other vertices in V. By applying the Chernoff bound from Theorem 21 we see that

$$\mathbb{P}(\alpha^{\mathrm{c}}) \leqslant e^{-\Theta(n)} \ ,$$

as required. ■

Note that we will later make use of the fact that when α occurs, the random vector $\mathbf{I}_u^{(3)}(A) - \mathbf{I}_u^{(3)}(A')$ has at least $\binom{np(1-p)/2}{3} = \Theta(n^3)$ non zero entries. This is the case since at least $\frac{np(1-p)}{2}$ of the vertices are without loss of generality contained in just A, so all entries corresponding to triples of these have value one in the vector $\mathbf{I}_u^{(3)}(A) - \mathbf{I}_u^{(3)}(A')$.

Lemma 47. *We have that* $\mathbb{P}(\beta) = 1 - e^{-\Theta(n)}$.

Proof. A fixed vertex $v \in V$ is connected to u in both A and A' with probability p^2 independently of any other vertices. Then

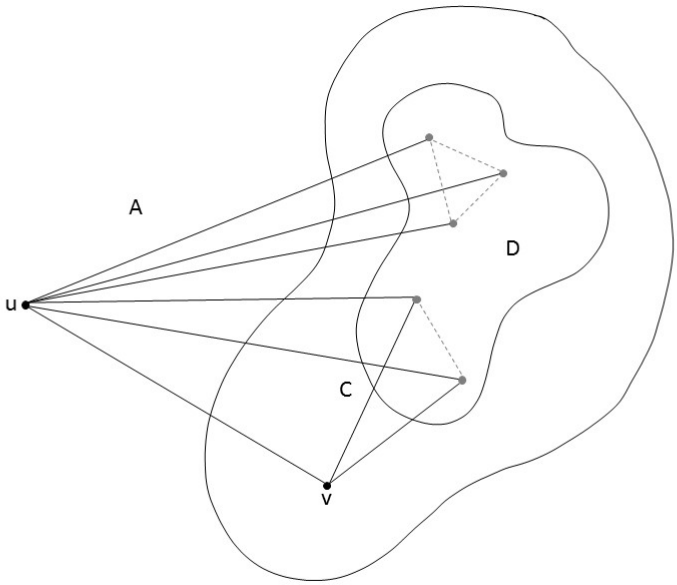
$$\begin{aligned} \mathbb{P}(\beta^{\mathrm{c}}) &= (1-p^2)^{n-1} \\ &= e^{-\Theta(n)} \ , \end{aligned}$$

as requried. ■

By applying a union bound to the complements of α and β we get the following bound the probability of η occurring.

Lemma 48. *We have that* $\mathbb{P}(\eta) \geq 1 - e^{-\Theta(n)}$.

From now on we assume that the event η, so (A, A') is "good", occurs. In the following we will now take A and A' to be fixed and condition on these.



Since we have assumed that η occurrs, we know that there exists an edge $e = \{u, v\}$ that is present in both A and A'. Define the vertex set W to be $[n] \setminus \{u, v\}$ and D to be the set of edges between vertices in W. Now define the vector $\mathbf{M}^{(3)}(D)$ to be the triangle adjacency vector of size $\binom{n-2}{3}$ of all vertices apart from u and v. In other words, each entry of $\mathbf{M}^{(3)}(D)$ corresponds to a set of size three of the vertices in W and the entry corresponding to the set of vertices $\{i, j, k\}$ is equal to one if all edges between i, j and k are present in G and zero otherwise. We define $\mathbf{I}_u^{(2)}(D)$ analogous to $\mathbf{I}_u^{(3)}(A)$, but for sets of vertices in W of size two. Similarly, we define $\mathbf{M}^{(2)}(D)$ analogous to $\mathbf{M}^{(3)}(D)$ for sets of vertices in W of size 2. We will denote by $*$ entry-wise multiplication of vectors. Then Ψ is equal to

$$\Psi = \left\langle \mathbf{I}_u^{(3)}(A) - \mathbf{I}_u^{(3)}(A'),\, \mathbf{M}^{(3)}(D) \right\rangle + \left\langle \left(\mathbf{I}_u^{(2)}(A) - \mathbf{I}_u^{(2)}(A')\right) * \mathbf{I}_v^{(2)}(C),\, \mathbf{M}^{(2)}(D) \right\rangle \ .$$

By applying Lemma 17 to $\mathbf{X} = \mathbb{E}_{C,D}[e^{i\theta\Psi}]$ we get

$$|\mathbf{X}|^2 \leqslant \mathbb{E}_{C,C'}\left[\mathbb{E}_D\left[e^{i\theta\left\langle \mathbf{R}_{u,v}(C,C'),\, \mathbf{M}^{(2)}(D)\right\rangle}\right]\right] \ , \tag{2.91}$$

where $\mathbf{R}_{u,v}(C, C') = \left(\mathbf{I}_u^{(2)}(A) - \mathbf{I}_u^{(2)}(A')\right) * \left(\mathbf{I}_v^{(2)}(C) - \mathbf{I}_v^{(2)}(C')\right)$.

To bound $\left|\mathbb{E}_B\left[e^{i\theta\left\langle \mathbf{I}_u^{(3)}(A) - \mathbf{I}_u^{(3)}(A'),\, \mathbf{M}^{(3)}(B)\right\rangle}\right]\right|$ it thus suffices to show that the expectation

$$\mathbf{Y} = \mathbb{E}_D\left[e^{i\theta\left\langle \mathbf{R}_{u,v}(C,C'),\, \mathbf{M}^{(2)}(D)\right\rangle}\right]$$

is small in absolute value.

We now say that the pair C, C' is "good" if there exists a set of size at least $np^2(1-p)^2$ vertices in the symmetric difference of both A and A' and C and C'. Let Γ be the event that C, C' is good.

Lemma 49. *We have that* $\mathbb{P}(\Gamma \mid \eta) \geq 1 - e^{-\Theta(n)}$.

Proof. Since we are conditioning on η, so the event that A, A' is good, we have that the symmetric difference between A and A' has size at least $np(1-p)$. Every endvertex in $D(A, A')$ lies in $D(C, C')$ with probability $2p(1-p)$ independently of any other vertices. By applying the Chernoff bound from Theorem 21 we get

$$\mathbb{P}(\Gamma \mid \eta) \geq 1 - e^{-\Theta(n)} \ ,$$

as required. ■

From now on, assume that both Γ and η occur. As a result, there is a set of at least $np^2(1-p)^2$ vertices in $D(A, A') \cap D(C, C')$. It follows that there exists a set of at least $\frac{np^2(1-p)^2}{4}$ vertices that are without loss of generality all contained in $D_A(A, A') \cap D_C(C, C')$. Denote this set of vertices by R. Furthermore, note that all entries of $\mathbf{R}_{u,v}(C, C')$ depending only on vertices in R have non-zero entries. So when Γ and η both occur, $\mathbf{R}_{u,v}(C, C')$ has at least $\Theta(n^2)$ non-zero entries. This allows us to get the following bound on $|\mathbf{Y}|$ for

$\theta \in (n^{-0.6}, \pi)$

$$
\begin{aligned}
|\mathbf{Y}| &= \left| \mathbb{E}_D \left[\prod_{e \in \binom{D}{2}} e^{i\theta (\mathbf{R}_{u,v}(C,C'))_e \cdot \left(\mathbf{M}^{(2)}(D)\right)_e} \right] \right| \\
&= \left| \mathbb{E}_D \left[\prod_{e \in \binom{D}{2},\, (\mathbf{R}_{u,v}(C,C'))_e \neq 0} e^{i\theta (\mathbf{R}_{u,v}(C,C'))_e \cdot \left(\mathbf{M}^{(2)}(D)\right)_e} \right] \right| \\
&= \left| \prod_{e \in \binom{D}{2},\, (\mathbf{R}_{u,v}(C,C'))_e \neq 0} \mathbb{E}_D \left[e^{i\theta (\mathbf{R}_{u,v}(C,C'))_e \cdot \left(\mathbf{M}^{(2)}(D)\right)_e} \right] \right| \text{ by independence of } \left(\mathbf{M}^{(2)}(D)\right)_e \\
&\leqslant \prod_{e \in \binom{D}{2},\, (\mathbf{R}_{u,v}(C,C'))_e \neq 0} \left(1 - 8p(1-p) \left|\left| \frac{\theta \cdot (\mathbf{R}_{u,v}(C,C'))_e}{2\pi} \right|\right|^2 \right) \text{ by Lemma 14} \\
&= \prod_{e \in \binom{D}{2},\, (\mathbf{R}_{u,v}(C,C'))_e \neq 0} \left(1 - 8p(1-p) \left(\frac{\theta \cdot (\mathbf{R}_{u,v}(C,C'))_e}{2\pi} \right)^2 \right) \\
&= \prod_{e \in \binom{D}{2},\, (\mathbf{R}_{u,v}(C,C'))_e \neq 0} \left(1 - 8p(1-p) \left(\frac{\theta}{2\pi} \right)^2 \right) \text{ as } |\mathbf{R}_{u,v}(C,C')|_e \in \{0, \pm 1\} \\
&\leqslant \left(1 - 8p(1-p) \left(\frac{\theta}{2\pi} \right)^2 \right)^{\Theta\left(n^2\right)} \text{ since } R_{u,v}(A;C) \text{ has at least } \Theta(n^2) \text{ non-zero entries} \\
&= e^{-8p(1-p)\left(\frac{\theta}{2\pi}\right)^2 \Theta\left(n^2\right)} \\
&= e^{-\Theta\left(\theta^2 n^2\right)} \ .
\end{aligned}
$$

Recall from (2.90) and (2.91) that

$$|\mathbb{E}[e^{i\theta \mathbf{S}_n}]| \leqslant \left(\mathbb{E}_{A,A',B}[|\mathbf{X}|]\right)^{1/2} \ .$$

Note that for a fixed 'good' (A, A'), the probability that (C, C') are also good is at least $1 - e^{-\Theta(n)}$ by Lemma 49. So for a fixed 'good' pair (A, A'), we can bound $\mathbb{E}[\mathbf{Y}]$ as follows:

$$
\begin{aligned}
|\mathbb{E}_{C,C'}[\mathbf{Y} \mid \eta]| &= \mathbb{P}(\Gamma \mid \eta) \cdot \mathbb{E}_{C,C'}[\mathbf{Y} \mid \eta, \Gamma] + \mathbb{P}(\Gamma^{\mathrm{c}} \mid \eta) \cdot \mathbb{E}_{C,C'}[\mathbf{Y} \mid \eta, \Gamma^{\mathrm{c}}] \\
&\leqslant e^{-\Theta\left(\theta^2 n^2\right)} + e^{-\Theta(n)} \ .
\end{aligned}
$$

So if we now also take into account the probability of A, A' being good, which is by Lemma 48 at least $1 - e^{\Theta(n)}$, we get the following bound on the absolute value of the

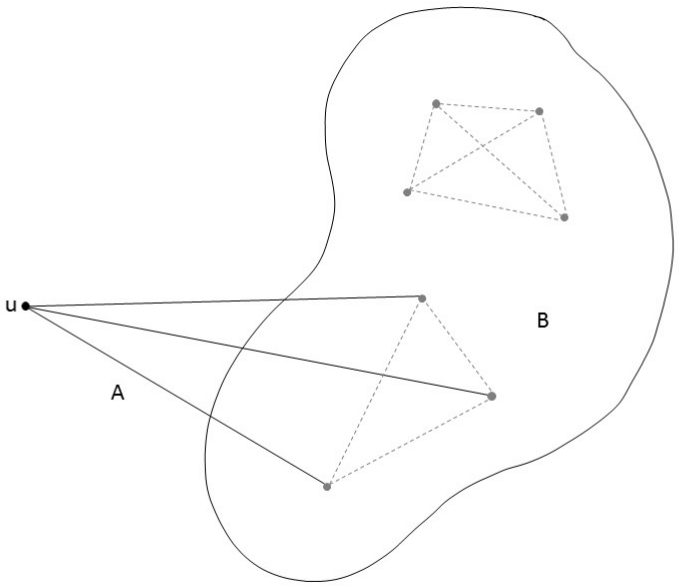
expectation of the characteristic function

$$\begin{aligned}\left|\mathbb{E}\left[e^{i\theta \mathbf{S}_n}\right]\right| &\leqslant \left(\left|\mathbb{E}_{A,A',B}[\mathbf{X}]\right|\right)^{\frac{1}{4}} \\ &\leqslant \left(\left|\mathbb{E}_{C,C'}[\mathbf{Y} \mid \eta]\right| \cdot \mathbb{P}(\eta) + \left|\mathbb{E}_{C,C'}[\mathbf{Y} \mid \eta^{\mathrm{c}}]\right| \cdot \mathbb{P}(\eta^{\mathrm{c}})\right)^{\frac{1}{4}} \\ &\leqslant \left(e^{-\Theta(n)} + e^{-\Theta(n)} + e^{-\Theta\left(\theta^2 n^2\right)}\right)^{1/4} \\ &\leqslant e^{-\Theta(n)} + e^{-\Theta(\theta^2 n^2)} \ .\end{aligned}$$

Since in region R_7 we have that $\theta \geqslant n^{-0.6}$, this is bounded above by $e^{-\Theta\left(n^{0.8}\right)}$ for all $\theta \in R_7$ as required. This proves Lemma 13.

2.7.7 Region $R_6 = \left(n^{-1.4}, n^{-0.6}\right)$

We now prove Lemma 12. For the first half of this section, the setup is the same as in the first half of the previous section. As before, we partition the vertex set $[n]$ into two sets, U and V, where U consists of a single vertex, u say. Denote by A the edges between u and V and B the edges inside V.



The vectors $\mathbf{I}_u^{(2)}(A)$, $\mathbf{M}^{(3)}(B)$ and $\mathbf{M}^{(2)}(B)$ are defined as in the previous section. As in equation (2.89), $\mathbf{S}_n$ is then equal to the number of K_4 inside G with one or zero vertices in U, so

$$\mathbf{S}_n = \left\langle \mathbf{I}_u^{(3)}(A), \mathbf{M}^{(3)}(B) \right\rangle + \left\langle \mathbf{M}^{(4)}(B), \mathbf{I} \right\rangle \ .$$

By applying Lemma 17 and the same argument as in the previous section, we see that to prove that $\mathbb{E}[e^{i\theta \mathbf{S}_n}]$ is small in absolute value, it suffices to show that for most choices of

A, A' the expectation

$$\mathbf{Y} = \left| \mathbb{E}_B \left[e^{i\theta \langle \mathbf{I}_u^{(3)}(A) - \mathbf{I}_u^{(3)}(A'), \mathbf{M}^{(3)}(B) \rangle} \right] \right|$$

is small in absolute value.

We define a pair A, A' to be "good" slightly differently in this section to the previous one. The event α is still defined to be the event that the symmetric difference between A and A' has size at least $np(1-p)$. The event β is defined, differently to the last section, to be the event that there exists a set of $\frac{np^2}{2}$ vertices in V that are all connected to u in both A and A'. The event that A, A' is "good", denoted by η, is now defined to be the event $\alpha \cap \beta$. To bound the probability of η occurring, we now bound the probabilities of α and β occurring separately. We define the sets $D(A, A')$ and $D_A(A, A')$ as in the previous section.

Lemma 50. *We have that* $\mathbb{P}(\alpha) \geq 1 - e^{-\Theta(n)}$.

Proof. As α is the same event as in the last section, the proof of Lemma 46 still applies. ■

We further have the following lemma.

Lemma 51. *We have that* $\mathbb{P}(\beta) \geq 1 - e^{-\Theta(n)}$.

Proof. For a vertex $v \in V$, the probability that v is connected to u in A and A' is equal to p^2 independently of any other vertices in V being connected to u. The number of $v \in V$ connected to u in both A and A', denoted by Ψ, is thus the sum of $|V| = n-1$ independent Bernoulli random variables each with success probability p^2. By the Chernoff bound from Theorem 21 then

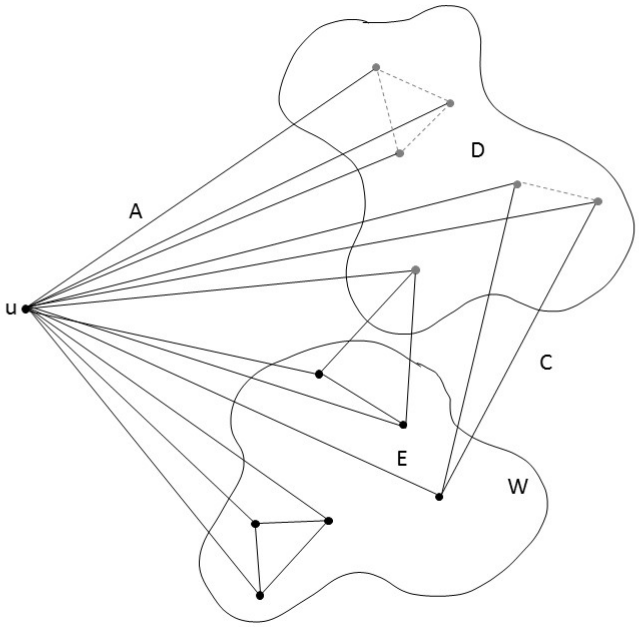
$$\mathbb{P}(\beta^{\mathrm{c}}) = \mathbb{P}\left(\Psi \leqslant \frac{np^2}{2}\right) \leqslant e^{-\Theta(n)} \tag{2.92}$$

as required. ■

We can now bound the probability of A, A' being good by applying a union bound to the complements.

Lemma 52. *We have that* $\mathbb{P}(\eta) \geq 1 - e^{-\Theta(n)}$.

From now on, we assume that the event η occurs and condition on the edges A, A', so take these as fixed. We now partition the vertex set further. As we have conditioned on η occurring, we know that there exists a set of $\frac{np^2}{2}$ vertices in V that are all adjacent to u in both A and A'. Denote this set by W. Further denote the set of edges between vertices in W by E, the set of edges between W and $V \setminus W$ by C, and the set of edges between vertices inside $V \setminus W$ by D.



Then $\mathbf{X} = \langle \mathbf{I}_u^{(3)}(A) - \mathbf{I}_u^{(3)}(A'),\, \mathbf{M}^{(3)}(B) \rangle$ is equal to the difference of the number of K_4 between A and A' with zero, one, two or three vertices inside W. Extending vectors by zeros if necessary, we thus have

$$\begin{aligned}
\mathbf{X} = {} & \langle \mathbf{I}_u^{(3)}(A) - \mathbf{I}_u^{(3)}(A'),\, \mathbf{M}^{(3)}(D) \rangle \\
& + \sum_{v \in W} \langle (\mathbf{I}_u^{(2)}(A) - \mathbf{I}_u^{(2)}(A')) * \mathbf{I}_v^{(2)}(C),\, \mathbf{M}^{(2)}(D) \rangle \\
& + \sum_{(v,v') \in E(W)} (\langle \mathbf{I}_u(A) * \mathbf{I}_v(C),\, \mathbf{I}_u(A) * \mathbf{I}_{v'}(C) \rangle - \langle \mathbf{I}_u(A') * \mathbf{I}_v(C),\, \mathbf{I}_u(A') * \mathbf{I}_{v'}(C) \rangle) \\
& + \langle \mathbf{I}_u^{(3)}(A) - \mathbf{I}_u^{(3)}(A'),\, \mathbf{M}^{(3)}(E) \rangle \ ,
\end{aligned}$$

where $\mathbf{I}_u(A)$ and $\mathbf{I}_u^{(2)}(A)$ are defined as in the previous section. For simplicity of notation, we further define the following vectors:

$$\begin{aligned}
\mathbf{P}_u &= \left(\mathbf{I}_u^{(3)}(A) - \mathbf{I}_u^{(3)}(A')\right) \\
\mathbf{R}_u &= \left(\mathbf{I}_u^{(2)}(A) - \mathbf{I}_u^{(2)}(A')\right) \\
\mathbf{A}_u(C,E) &= \sum_{(v,v') \in E(W)} (\langle \mathbf{I}_u(A) * \mathbf{I}_v(C),\, \mathbf{I}_u(A) * \mathbf{I}_{v'}(C) \rangle - \langle \mathbf{I}_u(A') * \mathbf{I}_v(C),\, \mathbf{I}_u(A') * \mathbf{I}_{v'}(C) \rangle) \ .
\end{aligned}$$

By applying Lemma 18 to $\mathbf{Y}$, we have

$$\begin{aligned}
|\mathbf{Y}|^2 &= \left| \mathbb{E}_{C,D,E} \left[e^{i\theta(\left\langle \mathbf{P}_u, \mathbf{M}^{(3)}(D) \right\rangle + \left\langle \mathbf{P}_u, \mathbf{M}^{(3)}(E) \right\rangle + \sum_{v \in W} \left\langle \mathbf{R}_u * \mathbf{I}_v^{(2)}(C), \mathbf{M}^{(2)}(D) \right\rangle + \mathbf{A}_{\{u\}}(C,E))} \right] \right|^2 \\
&\leqslant \mathbb{E}_{C,C',E} \left[\left| \mathbb{E}_D \left[e^{i\theta \sum_{v \in W} \left\langle \mathbf{R}_u * \left(\mathbf{I}_v^{(2)}(C) - \mathbf{I}_v^{(2)}(C')\right), \mathbf{M}^{(2)}(D) \right\rangle} \right] \right| \right]
\end{aligned}$$

and so to bound $\left|\mathbb{E}\left[e^{i\theta\mathbf{X}}\right]\right|$ it suffices to show that the expectation $\mathbf{Z} = \mathbb{E}_D\left[e^{i\theta\Xi}\right]$, where $\Xi = \sum_{v\in W}\left\langle \mathbf{R}_u * \left(\mathbf{I}_v^{(2)}(C) - \mathbf{I}_v^{(2)}(C')\right), \mathbf{M}^{(2)}(D)\right\rangle$, is small in absolute value for most choices of C, C' and E. To simplify notation, for an edge $e \in \binom{K}{2}$ we let $T_e = \left(\sum_{v\in W}\left(\mathbf{I}_v^{(2)}(C) - \mathbf{I}_v^{(2)}(C')\right)\right)_e$.

Since A, A' is "good", we have that the symmetric difference of A and A' has size at least $np(1-p)$. We now wish to show that at least a constant fraction of the endvertices d of edges in $D(A, A') \subseteq V \setminus W$ are connected to around $\sqrt{|W|}$ vertices $w \in W$ such that w is a common neighbour of d in exactly one of C and C'. In other words, we wish to show that for at least a constant fraction of the edges $e \in \binom{D(A,A')}{2}$, T_e is of size roughly $\sqrt{|W|}$. Note that $np(1-p) \leqslant D(A, A') \leqslant n$ and that for each $e \in \binom{D(A,A')}{2}$, the distribution of T_e is the difference of two binomials of the form $\mathrm{Bin}(|W|, p^2)$. We should hence typically have T_e around $\sqrt{|W|}$ in magnitude. We now go on to prove this rigorously.

Since we have assumed that η, so A, A' is "good" occurs, we know that the symmetric difference of A and A' has size at least $np(1-p)$. We now wish to show that at least a constant fraction of the edges between vertices in $D(A, A')$ also satisfy $|T_e| \in (|W|^{0.49}, |W|^{0.51})$ and call a pair C, C' "good" if it satisfies this.

We now define Λ_e for each $e \in \binom{D(A,A')}{2}$ to be the good event depending on C, C' that $|T_e| \in (|W|^{0.49}, |W|^{0.51})$ and let $\gamma = \mathbb{P}(\Lambda_e)$. By standard concentration and anti-concentration estimates for Binomial distributions, we have $\gamma \geqslant 0.1$ provided n is sufficiently large, depending on p. Then let Λ be the good event that for more than $\frac{1}{2}$ choices of the edges $e \in \binom{D(A,A')}{2}$, the event Λ_e occurs.

Lemma 53. *[14][Lemma 12] There is a constant A such that for every k*

$$\mathbb{P}(\Lambda) \geqslant 1 - \frac{k^{Ak}}{|K|^k}$$

and so by choosing $k = 200$ we get

$$\mathbb{P}(\Lambda) \geqslant 1 - \frac{1}{t^{100}} \ .$$

Proof. The proof of this lemma is essentially the same as the proof of Lemma 12 in [14]. ■

We now suppose that both Λ and η occur. For simplicity of notation, denote the vector $\sum_{v\in W}(R(A_u) - R(A'_u)) * (R(C_v) - R(C'_v))$ by $\mathbf{N}_{\{u,W\}}(A; C)$ in what follows. We

have

$$
\begin{aligned}
|\mathbf{Z}| &= \left| \mathbb{E}_D \left[\prod_{e \in \binom{W}{2}} e^{i\theta \left(\mathbf{N}_{\{u,W\}}(A;C)\right)_e \cdot \left(\mathbf{M}^{(2)}(D)\right)_e} \right] \right| \\
&= \left| \mathbb{E}_D \left[\prod_{e \in \binom{D(A,A')}{2}, \left(\mathbf{N}_{\{u,W\}}(A;C)\right)_e \neq 0} e^{i\theta \left(\mathbf{N}_{\{u,W\}}(A;C)\right)_e \cdot \left(\mathbf{M}^{(2)}(D)\right)_e} \right] \right| \\
&= \left| \prod_{e \in \binom{D(A,A')}{2}, \left(\mathbf{N}_{\{u,W\}}(A;C)\right)_e \neq 0} \mathbb{E}_D \left[e^{i\theta \left(\mathbf{N}_{\{u,W\}}(A;C)\right)_e \cdot \left(\mathbf{M}^{(2)}(D)\right)_e} \right] \right| \text{ by independence of } \left(\mathbf{M}^{(2)}(D)\right)_e \\
&= \prod_{e \in \binom{D(A,A')}{2}, \left(\mathbf{N}_{\{u,W\}}(A;C)\right)_e \neq 0} \left| 1 - 8p(1-p) \left|\left| \frac{\theta \cdot \left(\mathbf{N}_{\{u,W\}}(A;C)\right)_e}{2\pi} \right|\right|^2 \right| \text{ by Lemma 14} \\
&\leqslant \prod_{e \in \binom{D(A,A')}{2}, \neg\Lambda_e} \left| 1 - 8p(1-p) \left|\left| \frac{\theta \cdot \left(\mathbf{N}_{\{u,W\}}(A;C)\right)_e}{2\pi} \right|\right|^2 \right| \\
&= \prod_{e \in \binom{D(A,A')}{2}, \neg\Lambda_e} \left(1 - 8p(1-p) \left(\frac{\theta \cdot \left(\mathbf{N}_{\{u,W\}}(A;C)\right)_e}{2\pi} \right)^2 \right) \\
&\qquad \text{since } \theta = o\left(n^{-0.51}\right), |\left(\mathbf{N}_{\{u,W\}}(A;C)\right)_e| < |W|^{0.51} \\
&= \prod_{e \in \binom{D(A,A')}{2}, \neg\Lambda_e} \left(1 - 8p(1-p) \left(\frac{\theta \cdot |D(A,A')|^{0.49}}{2\pi} \right)^2 \right) \text{ since } |\left(\mathbf{N}_{\{u,W\}}(A;C)\right)_e| \geq |W|^{0.49} \\
&\leqslant e^{-8p(1-p)\left(\frac{\theta}{2\pi} \cdot \left(\frac{np^2}{2}\right)^{0.49}\right)^2 \frac{|D(A,A')|^2}{8}} \text{ since } \Lambda \text{ occurs} \\
&= e^{-\Theta\left(\theta^2 \cdot n^{2.98}\right)} \text{ since } |D(A,A')| \geq np(1-p) \ .
\end{aligned}
$$

If we also take into account the probabilities of η and Λ, we get the following bound on the absolute value of the expectation of the characteristic function

$$
\begin{aligned}
\left|\mathbb{E}\left[e^{i\theta\mathbf{S}}\right]\right| &\leqslant \left(e^{-\Theta(n)} + \frac{1}{t^{100}} + e^{-\Theta\left(\theta^2 \cdot n^{2.98}\right)} \right)^{\frac{1}{4}} \\
&\leqslant \frac{1}{t^{25}} + e^{-\Theta\left(\theta^2 \cdot n^{2.98}\right)} \ .
\end{aligned}
$$

Since $\theta \geq n^{-1.4}$ in region R_6, this implies that in region R_6 we have $\left|\mathbb{E}\left[e^{i\theta\mathbf{S}}\right]\right| \leqslant t^{-25}$ as required. This proves Lemma 12.

2.8 Conclusion and Further Open Problems

We have thus managed to prove Lemmas 7 - 13, which is sufficient to prove our main result. This means we have established a local limit theorem for the number of K_4 subgraphs in the random graph $G(n,p)$. Naturally, the question whether this can be generalised to K_k for any fixed integer k arises. Looking at our conditioning arguments, one can see that they are in fact similar to the conditioning arguments used in the Local limit theorem for triangle count paper by Gilmer and Kopparty [14]. In their paper, they have two main conditioning arguments: One of them conditioning on a set of edges outside a single or multiple matchings, while the other splits the vertices into two sets and conditions on the edges inside one of the sets. In this chapter, we have applied similar conditioning arguments to these two main arguments, but applied them in stages. For example, in regions R_3 and R_2 we have again conditioned on everything outset subsets of edges in matchings in a more generalised way. In Regions R_5 and R_4, we first split the vertices into two sets U and V and condition on all edges inside one of the sets U. We then go on to apply the other conditioning argument, conditioning on everything outside a set of matchings, to the set V. In the final two regions, R_7 and R_6 we then apply the conditioning argument that splits the vertices into two sets and conditions on the edges inside one of the sets twice. It thus appears, that this method of using layers of the same conditioning arguments should work in a more generalised setting and would allow the proof of a local central limit theorem for the number of K_k in $G(n,p)$ for any fixed k and $p \in (0,1)$. Note that since writing the first draft of this, Berkowitz [2] has managed to prove something similar to this, but in a different way, proving the result for all K_k.

Another question that the local central limit theorem for the number of triangles and K_4 in $G(n,p)$ for fixed p brings up is the question whether a similar result can be shown for the number of triangles in $G(n,p)$ throughout a range of non-constant p, where this makes sense (for it to make sense, we certainly need $np \to \infty$ as $n \to \infty$ to ensure the expected number of triangles goes to infinity as well as $n^2(1-p) \to \infty$ as $n \to \infty$ to ensure that the expected number of missing triangles goes to infinity). Most of the arguments we have used do give good enough bounds for sufficiently close to constant p, so we assume that our arguments will get close to proving a local limit theorem for non-constant p also. In particular, some of the refinements we have made to the arguments used in [14] should allow for better bounds on the characteristic function in the triangle case. It is yet to be seen whether these bounds will be sufficiently good for all ranges of t.

Chapter 3

Graphs on Non-constant Surfaces

3.1 Introduction

For a function $g(n)$ taking values in the non-negative integers, let $\mathscr{G}_n^{g(n)}$ be the class of graphs on n vertices embeddable in $\mathbf{S}_{\lfloor g(n)/2 \rfloor}$, the orientable surface of Euler genus $2 \cdot \lfloor g(n)/2 \rfloor$. Similarly, we define $\mathscr{H}_n^{g(n)}$ to be the class of graphs on n vertices embeddable in $\mathbf{N}_{\lfloor g(n) \rfloor}$, the non-orientable surface of Euler genus $\lfloor g(n) \rfloor$. In the following, we will investigate the growth with respect to n of the sizes of these graph classes for different non-negative functions $g(n)$ and which properties a typical graph in these classes displays.

In what follows, all graphs considered will be finite, simple and labeled unless otherwise specified. For a class of graphs $\mathscr{A}$, we denote by $\mathscr{A}_n$ the class of all graphs contained in $\mathscr{A}$ on the vertex set $[n] = \{1, \ldots, n\}$. To simplify notation, we also denote by $R \in_U \mathscr{A}$ a graph chosen uniformly at random from all graphs in the class $\mathscr{A}$ and by $R_n \in_U \mathscr{A}_n$ a graph chosen uniformly at random from all graphs on n vertices in the class $\mathscr{A}$. Before we start considering graphs on surfaces, we define what we mean by orientable and non-orientable surfaces of Euler genus $g(n)$.

Definition. *[48] Let* $\mathbf{S}$ *be the surface obtained from the sphere by adding* h *handles and* c *crosscaps. If* $c = 0$*, then* $\mathbf{S} = \mathbf{S}_h$*, the* ***orientable surface of Euler genus 2h****. Otherwise* $\mathbf{S} = \mathbf{N}_{2h+c}$*, the* ***non-orientable surface of Euler genus 2h+c****.*

It can then be shown that every surface is homeomorphic, written as $\sim$, to one of $\mathbf{S}_h$ $(h \geqslant 0)$ or $\mathbf{N}_k$ $(k \geqslant 1)$:

Theorem 54. *[54]* ***(Classification of Surfaces)*** *Every surface is homeomorphic to precisely one of the surfaces* $\mathbf{S}_h$ *(*$h \geqslant 0$*) or* $\mathbf{N}_k$ *(*$k \geqslant 1$*).*

We thus define the Euler genus of a surface as follows:

Definition. *The* ***Euler genus*** *of a surface* $\mathbf{S}$ *is defined to be equal to* $2h$ *if* $\mathbf{S} \sim \mathbf{S}_h$ *and equal to* k *if* $\mathbf{S} \sim \mathbf{N}_k$*.*

This is well defined by Theorem 54. In the following we will usually treat the orientable and non-orientable cases separately.

We call a graph G *planar*, if there exists a drawing of G in the plane, or equivalently the sphere, such that no two edges intersect other than at their endpoints. Similarly, we define a graph G to be embeddable in a surface $\mathbf{S}$ if there exists a drawing of G in $\mathbf{S}$ such that no two edges intersect. More formally, a graph G is embedded in a topological space $\mathbf{S}$ if the vertices of G are distinct elements of $\mathbf{S}$ and every edge of G is a simple arc connecting in $\mathbf{S}$ the two vertices which it joins in G, such that its interior is disjoint from other edges and vertices [48]. It is important to note that we will not be focusing on embeddings of graphs in surfaces, but on graphs G for which an embedding of G in $\mathbf{S}$ exists. An embedding of a connected graph G in a surface $\mathbf{S}$ such that all faces are open 2-cells, that is homeomorphic to an open rectangle in $\mathbb{R}^2$, is called a cellular embedding and in this case we say that G is cellularly embeddable in $\mathbf{S}$.

Definition. *We define the* ***orientable Euler genus*** *of a graph G to be the least integer $2h$ such that G has an embedding in the surface $\mathbf{S}_h$. Similarly, the* ***non-orientable Euler genus*** *of a graph G is the least integer h such that G has an embedding in the surface $\mathbf{N}_h$. Furthermore, the* ***Euler genus*** *of a graph G is defined to be the minimum of its orientable and non-orientable Euler genus.*

Definition. *We define $\mathscr{P}$ to be the class of all planar graphs. Furthermore, for some fixed constant h we define $\mathscr{G}^h$ and $\mathscr{H}^h$ to be the classes of all graphs embeddable in $\mathbf{S}_{h/2}$ or $\mathbf{N}_h$ respectively. We further define $\mathscr{H}^0 = \mathscr{G}^0$. Similarly, given a non-negative function $g(n)$, for each n we define $\mathscr{G}_n^{g(n)}$ and $\mathscr{H}_n^{g(n)}$ to be the classes of all labelled graphs on the vertex set $[n]$ embeddable in the surface $\mathbf{S}_{g(n)/2}$ and $\mathbf{N}_{g(n)}$ respectively, that is for example $\mathscr{G}_n^{g(n)} = \left(\mathscr{G}^{g(n)}\right)_n$. The classes $\mathscr{G}^{g(n)}$ and $\mathscr{H}^{g(n)}$ are then defined as follows:*

$$\mathscr{G}^g = \bigcup_{n=1}^{\infty} \mathscr{G}_n^{g(n)}$$
$$\mathscr{H}^g = \bigcup_{n=1}^{\infty} \mathscr{H}_n^{g(n)} \ .$$

When considering the classes $\mathscr{G}_n^{g(n)}$ and $\mathscr{H}_n^{g(n)}$ we usually implicitly mean $g(n)$ to be $\lfloor g(n) \rfloor$. In what follows, we will sometimes suppress the dependence on n when talking about genus functions, that is write g instead of $g(n)$, to avoid cluttering up formulas. Note that by their definitions all of these classes are closed under isomorphism if we look at the classes on n vertices rather than the vertex set $[n]$. One of our main objectives is to understand the growth of the sizes of $\mathscr{G}_n^{g(n)}$ and $\mathscr{H}_n^{g(n)}$ for large n and different functions $g(n)$. Note that it is not worth considering $g(n)$ growing faster than the maximum genus

a graph on n vertices can have, which is the orientable or non-orientable Euler genus of the complete graph on n vertices. By [51], the orientable Euler genus $\mathbf{g}^o$ of a complete graph on $n \geqslant 3$ vertices is equal to

$$\mathbf{g}^o(K_n) = 2 \cdot \left\lceil \frac{(n-3)(n-4)}{12} \right\rceil \ . \tag{3.1}$$

Similarly, by [50], the non-orientable Euler genus $\mathbf{g}^n$ of a complete graph on $n \geqslant 3$ vertices is equal to

$$\mathbf{g}^n(K_n) = \left\lceil \frac{(n-3)(n-4)}{6} \right\rceil \ , \tag{3.2}$$

except when $n = 7$ in which case $\mathbf{g}^n(K_7) = 3$. Hence the orientable Euler genus of K_n is always at least the non-orientable Euler genus of K_n except when $n = 7$. To help understand the growth of these graph classes, we define what it means for a class to be small, have a growth constant or be over-large. For this, we consider the radius of convergence of the exponential generating function of labeled graphs, which is for a class of graphs $\mathscr{A}$ defined to be

$$A(x) = \sum_{n \geqslant 0} \frac{|\mathscr{A}_n|}{n!} \cdot x^n \ . \tag{3.3}$$

This allows us to define the following measurements of the size of a graph class.

Definition. *We say that a class of graphs $\mathscr{A}$ is **small** if the radius of convergence of its exponential generating function is greater than zero. This is equivalent to saying that there exists a constant $c > 0$ and an $N \in \mathbb{N}$ such that for all $n \geqslant N$*

$$\left(\frac{|\mathscr{A}_n|}{n!} \right)^{\frac{1}{n}} < c \ .$$

Definition. *We say that a class of graphs $\mathscr{A}$ has **growth constant** α, where $0 < \alpha < \infty$ if*

$$\left(\frac{|\mathscr{A}_n|}{n!} \right)^{\frac{1}{n}} \to \alpha$$

as n tends to infinity.

It is worth noting, that a class of graphs $\mathscr{A}$ which has growth constant γ has radius of convergence of the exponential generating function equal to $\rho = \frac{1}{\gamma}$. The reverse implication does however not necessarily hold.

Definition. *We say that a class of graphs $\mathscr{A}$ is **over-large** if the radius of convergence of its exponential generating function is equal to zero. In other words, a class of graphs $\mathscr{A}$ is over-large if for every constant $c > 0$ and for every $N \in \mathbb{N}$ there exists an $n \geqslant N$ such that*

$$\left(\frac{|\mathscr{A}_n|}{n!} \right)^{\frac{1}{n}} > c \ .$$

The other main objective in what follows will be to find typical properties of a graph chosen uniformly at random from one of the before mentioned graph classes. Note that some of these properties can be readily deduced from the graph class being small, having a growth constant or being over-large, so our two main objectives are intertwined.

Some other properties can be deduced from the graph classes having some special properties, which we will define now. The first property we will define is the pendant appearance of a graph. We will later show that if a large proportion of graphs in a graph class have linearly many pendant vertices, one can deduce results concerning for example the maximum degree or the maximum face size of most graphs in the class. It turns out that weaker conclusions can still be drawn if the number of pendant vertices is sub-linear but still relatively large.

Definition. *Let G be a graph on the vertex set $\{1, \dots, n\}$. We say that a graph H on the vertex set $\{1, \dots, h\}$ for $h < n$ has a* ***pendant appearance*** *at W in G, where W is a subset of the vertices of G of size h, if the increasing bijection from $\{1, \dots, h\}$ to W gives an isomorphism between H and the induced subgraph $G[W]$; and there is only one edge between $G[W]$ and the rest of the graph G, and it is incident to the least element of W.*

For a class of graphs $\mathscr{A}$, we say that a rooted graph H with root vertex r ***can be attached*** *to $\mathscr{A}$ if for each graph G in $\mathscr{A}$ and any vertex v in G, the graph obtained by taking vertex disjoint copies of H and G and adding an edge between r and v, is still contained in the class $\mathscr{A}$.*

We now define some further properties of graph classes. Assume that a planar graph G has more than one component. We can then add an edge between any two vertices u and v in distinct components of G without the graph losing its planarity property. This is since we can choose an embedding of G in the plane in which both u and v are incident to the outer face and adding an edge e between u and v in this embedding does not cause any edge intersections, so $G \cup \{e\}$ is still a planar graph. This property is called bridge-addability. Furthermore, a graph is planar if and only if all of its components are planar. This property is called decomposability. A class is called addable if it is both decomposable and bridge-addable. We define these formally as follows:

Definition. *A non-empty finite class of graphs $\mathscr{A}$ is defined to be* ***bridge-addable*** *if for each graph G in $\mathscr{A}$, if u and v are vertices in distinct components of G, then the graph obtained from G by adding an edge joining u and v is also in $\mathscr{A}$. Note that in some places the term weakly addable is used instead of bridge-addable, with exactly the same meaning.*

A non-empty finite class of graphs $\mathscr{A}$ is defined to be ***decomposable*** *if for each graph*

G, the graph G is in $\mathscr{A}$ if and only if each component of G is in $\mathscr{A}$.
A non-empty finite class of graphs $\mathscr{A}$ is defined to be ***addable*** *if it is both bridge-addable and decomposable.*

We have already shown above that planar graphs are both bridge-addable and decomposable. Note however that the graph classes $\mathscr{G}^{g(n)}$ or $\mathscr{H}^{g(n)}$ in general do not satisfy the decomposability criterion, indeed they do so only in the extreme cases.

Proposition 55. *Let $\mathscr{A}_n = \mathscr{G}_n^{g(n)}$ or $\mathscr{A}_n = \mathscr{H}_n^{g(n)}$. Let $g : \mathbb{N} \to \mathbb{Z}_{\geqslant 0}$. Then, if $\mathscr{A}$ is not just $\mathscr{P}$ and is not all graphs, then $\mathscr{A}$ is not decomposable.*

Proof. Consider first $\mathscr{A} = \mathscr{G}^{g(n)}$. We assume that $g(n) \geqslant 0$ and even for each n; and that there exists an $a \in \mathbb{N}$ such that $a \geqslant 5$ with $g(a) \geqslant 2$, which we can do since $\mathscr{A}$ is not just $\mathscr{P}$, and $b \in \mathbb{N}$ with $g(b) < g^o(K_b)$, which we can do since $\mathscr{A}$ is not all graphs, where g^o denotes the orientable Euler genus in this case.

There are two cases. Suppose first that $g(n) \leqslant g^o(K_b)$ for each $n > b$. Then there exists $c \in \mathbb{N}$ such that $g(n) \leqslant c$ for each $n \in \mathbb{N}$. Since $g(a) \geqslant 2$, there is a connected non-planar graph H on $[a]$ in $\mathscr{A}$. Then $g^o((c+1)H) \geqslant g((c+1)H) > c$, where $(c+1)H$ denotes the graph consisting of $c+1$ copies of H, so $(c+1)H$ is not in $\mathscr{A}$; and so $\mathscr{A}$ is not decomposable.

Suppose instead that there exists $n > b$ such that $g(n) > g^o(K_b)$. Let H on $[n]$ be formed from K_b and a disjoint planar graph. Then H is in $\mathscr{A}$ since $g^o(H) = g^o(K_b) \leqslant g(n)$. But the component K_b of H is not in $\mathscr{A}$, since $g^o(K_b) > g(b)$. It follows again that $\mathscr{A}$ is not decomposable.

Now consider $\mathscr{A} = \mathscr{H}^{g(n)}$, where the proof is very similar. We have $g(n) \geq 0$ and integral for each n; and there exists an $a \in \mathbb{N}$ such that $a \geqslant 5$ with $g(a) \geq 1$, and $b \in \mathbb{N}$ with $g(b) < \tilde{g}(K_b)$, where $\tilde{g}$ denotes the non-orientable Euler genus in this case.

There are two cases. Suppose first that $g(n) \leqslant \tilde{g}(K_b)$ for each $n > b$. Then there exists $c \in \mathbb{N}$ such that $g(n) \leqslant c$ for each $n \in \mathbb{N}$. Since $g(a) \geqslant 1$, there is a connected non-planar graph H on $[a]$ in $\mathscr{A}$. Then $\tilde{g}((c+1)H) \geqslant g((c+1)H) > c$ so $(c+1)H$ is not in $\mathscr{A}$, where $(c+1)H$ denotes the graph consisting of $c+1$ copies of H; and so $\mathscr{A}$ is not decomposable.

Suppose instead that there exists $n > b$ such that $g(n) > \tilde{g}(K_b)$. Let H on $[n]$ be formed from K_b and a disjoint planar graph. Then H is in $\mathscr{A}$ since $\tilde{g}(H) = \tilde{g}(K_b) \leqslant g(n)$. But the component K_b of H is not in $\mathscr{A}$, since $\tilde{g}(K_b) > g(b)$. Thus again $\mathscr{A}$ is not decomposable. ■

Bridge-addability however is still maintained when moving from the plane to any surface or non-constant surfaces. These properties allow us to deduce some common properties of the graphs in the class, such as the following theorem about the proportion of connected graphs in a bridge-addable class. Before stating the theorem, we first need to make a few further definitions.

Definition. *For a graph G, we define $\kappa(G)$ to be the number of components in G. Further, we define* $\mathbf{Big}(\mathbf{G})$ *to be the subgraph of G induced by all the vertices in the largest component, where in the case of ties we choose the lexicographic first largest component. Finally, we define the fragment* $\mathbf{Frag}(\mathbf{G})$ *to be the subgraph of G induced by all the vertices not in the largest component. The number of vertices contained in the fragment is denoted by* $\mathbf{frag}(\mathbf{G})$.

It turns out that for all finite bridge-addable classes of graphs $\mathscr{A}$, a non-vanishing fraction of all graphs in $\mathscr{A}_n$ are connected.

Theorem 56. *[46, Theorem 2.2] Let $\mathscr{A}$ be any finite non-empty bridge-addable class of graphs and let $R \in_U \mathscr{A}$, so R is a graph sampled uniformly at random from $\mathscr{A}$. Then*

- *the random number $\kappa(R)$ of components of R is stochastically dominated by $1 + X$ where $X \sim Po(1)$, that is X has the Poisson distribution with mean 1*
- $\mathbb{P}(R \text{ is connected}) \geqslant \frac{1}{e}$
- $\mathbb{P}(\kappa(R) \geqslant k+1) \leqslant \frac{1}{k!}$
- $\mathbb{E}[\kappa(R)] \leqslant 2$.

Observe that for any non-negative function $g(n)$ and for any n, both $\mathscr{G}_n^{g(n)}$ and $\mathscr{H}_n^{g(n)}$ are bridge-addable, so Theorem 56 applies. When a growth constant exists, more precise results can be made about the probability of connectedness. It turns out that for all classes of graphs with a growth constant, the fraction of graphs that are connected is also bounded away from zero in the limit as n tends to infinity.

Theorem 57. *Let $\mathscr{A}$ be a class of graphs $\mathscr{G}^{g(n)}$ or $\mathscr{H}^{g(n)}$ which has a growth constant. Then for $R_n \in_U \mathscr{A}$ and for any fixed planar graph H,*

$$\liminf_{n\to\infty} \mathbb{P}(Frag(R_n) \cong H) > 0$$

and thus $\limsup_{n\to\infty} \mathbb{P}(R_n \text{ is connected}) < 1$.

In special cases, we can even deduce from bridge-addability or addability something about the size of the graph class. It turns out that a graph class that is both small and addable is actually guaranteed to have a growth constant.

Theorem 58. *[45, Theorem 3.3] Let $\mathscr{A}$ be a non-empty, small, addable class of finite graphs. Then there exists a constant $0 < c < \infty$ such that*

$$\left(\frac{|\mathscr{A}_n|}{n!}\right)^{\frac{1}{n}} \to c \ .$$

Sketch Proof. We define

$$a(n) = \frac{|\mathscr{A}_n|}{e^2 \cdot n!} \ .$$

Note that a superadditive function $b(n)$ is defined to be a function $b : \mathbb{N} \to \mathbb{R}_{\geqslant 0}$ such that for some n_0 and for all $i, j \in \mathbb{N}$

$$b(i + j) \geqslant b(i) \cdot b(j)$$

and

$$b(n) > 0 \text{ for all } n \geqslant n_0 \ .$$

By [52, Lemma 11.6] any superadditive function $b(n)$ satisfies

$$b(n)^{\frac{1}{n}} \to \sup_{n \in \mathbb{N}} \left\{ b(n)^{\frac{1}{n}} \right\} \ .$$

Since we assume that $\mathscr{A}$ is small, we know that $\sup_n \left\{ a(n)^{\frac{1}{n}} \right\} < \infty$. To prove the theorem it is hence sufficient to show that $a(n)$ is superadditive, as $\mathscr{A}$ being small then implies that the limit is strictly smaller than infinity, so $\mathscr{A}$ has a growth constant.

To show that $a(n)$ is superadditive, we note that by the decomposability of $\mathscr{A}$ the number of graphs in $\mathscr{A}$ is at least the number of ways we can construct a graph on two connected components, one with i and the other with j vertices, and such that each component is in $\mathscr{A}$. To avoid having to divide by two to account for the case $i = j$, we may just add any edge between the two components in this case, which we can do since the class is also bridge-addable. We furthermore have from Theorem 56 that at least $1/e$ of all graphs in $\mathscr{A}$ are connected, so for all $i, j \in [n]$ such that $i + j = n$ we have

$$|\mathscr{A}_n| \geqslant \binom{n}{i} \cdot \frac{|\mathscr{A}_i|}{e} \frac{|\mathscr{A}_j|}{e} \ . \tag{3.4}$$

Rearranging, we see that $a(n)$ is super-additive as required. ■

Using combinatorial arguments, we can make some statements about common properties in graph classes that have a growth constant. Probably the most important result of these, from which many other results follow, is the so called Pendant Appearances Theorem. It implies that the probability of observing fewer than linearly many pendant appearances of any fixed planar graph H in a random planar graph on n vertices is exponentially small. Note that the only condition of the theorem is that the class of graphs has a growth constant.

Theorem 59. *[46, Theorem 5.1]**(Pendant Appearances Theorem)** Let the class $\mathscr{A}$ of graphs have a (finite positive) growth constant c, and let $R_n \in_U \mathscr{A}_n$. Let H be a fixed graph with a distinguished 'root' vertex r, and suppose that H can be attached to $\mathscr{A}$ in the previously defined sense. Then there exists a constant $\alpha > 0$ such that*

$$\mathbb{P}(R_n \text{ has } \leqslant \alpha n \text{ pendant appearances of } H) = e^{-\Omega(n)} \ . \tag{3.5}$$

Sketch Proof. We define

$$\alpha = \frac{1}{9e^2 \cdot c^h \cdot (h+2) \cdot h!}$$

and aim to show that $\mathbb{P}(f_H(R_n) < \alpha n) \leqslant e^{-\alpha n}$ where $f_H(G)$ is the number of pendant appearances of H, a fixed graph on h vertices, in the graph G. Since the class $\mathscr{A}$ has a growth constant c, there exists an $n_0 \in \mathbb{N}$ such that $f(n) = |\mathscr{A}_n|$ satisfies for $n \geqslant n_0$

$$(1-\epsilon)^n \cdot n! \cdot c^n \leqslant f(n) \leqslant (1+\epsilon)^n \cdot n! \cdot c^n \ .$$

We now assume for a contradiction that $\mathbb{P}(f_H(R_n) \leqslant \alpha n) \geqslant e^{-\alpha n}$ and show that if this holds we can construct from graphs in $\mathscr{A}_n$ too many graphs in $\mathscr{A}_{(1+\delta)n}$ for $\mathscr{A}$ to have growth constant c, where $\delta = \alpha h$.

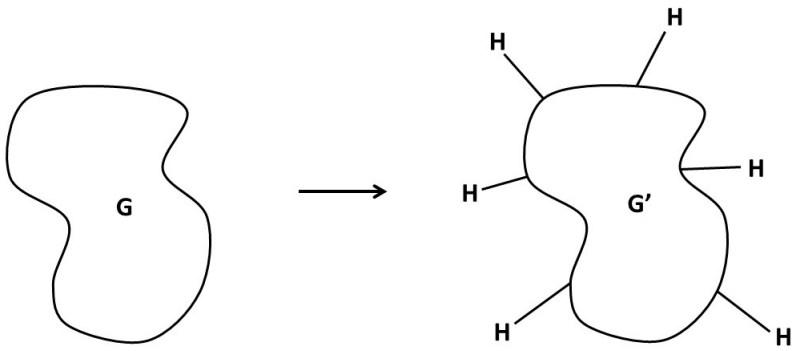


Figure 3.1: $G \in \mathscr{A}_n$ to $G' \in \mathscr{A}_{(1+\delta)n}$

We construct many graphs in $\mathscr{A}_{(1+\delta)n}$ from a graph G in $\mathscr{A}_n$ with few pendant copies of H as follows: Choose any subset of size δn from $[(1+\delta)n]$. These will be the so called "special" vertices in the graph that we construct. We put αn disjoint copies of H on these vertices. On the remaining n vertices, we put any graph $G \in \mathscr{A}$ with fewer than αn pendant appearances of H. Choosing graphs with fewer than αn pendant vertices for this will ensure that each graph is not constructed too many times. We then connect each of the copies of H on the "special" vertices to G by a single edge each incident to the root

of the copy of H. Since G has few pendant appearances of H, there are not too many pendant appearances of H in each of the constructions, thus implying that each graph is only constructed relatively few times. This suffices to show that we could construct more than $(1+\epsilon)^{(1+\delta)n} \cdot ((1+\delta)n)! \cdot c^{(1+\delta)n}$ graphs in $\mathscr{A}$ on $(1+\delta)n$ vertices if the probability of having many pendant copies of H is not sufficiently close to one. ■

From this result it has been possible to establish many bounds on properties of graphs chosen uniformly at random from a graph class with a growth constant. We will often use the terminology with high probability, which we will define now.

Definition. *We say that some property $\mathcal{P}$ holds for graphs in a graph class $\mathscr{A}$* ***with high probability*** *if it holds for a graph chosen uniformly at random from $\mathscr{A}_n$ with probability tending to one as n tends to infinity.*

Using the Pendant Appearances Theorem (see Theorem 59), it was possible to obtain bounds on the maximum degree of a graph chosen uniformly at random from a class of graphs with a growth constant. The maximum degree of a graph chosen uniformly at random from one of these classes is with high probability relatively small.

Theorem 60. *[43, Theorem 1.1] Let $\mathscr{A}$ be $\mathscr{G}^{g(n)}$ or $\mathscr{H}^{g(n)}$ and assume that $\mathscr{A}$ has a growth constant. Then there are constants $0 < c < C$ such that with high probability the maximum degree of a random graph $R_n \in_U \mathscr{A}_n$ satisfies*

$$c \log(n) \leq \Delta(R_n) \leq C \log(n) \ ,$$

where $\Delta(R_n)$ is the maximum degree of R_n.

Sketch Proof. We restrict ourselves to outlining the proof for the upper bound, since this is the part of the theorem we will repeatedly make use of later on. For the proof we will need Theorem 59 and the following lemma, which can be proven by a double counting argument.

Lemma 61. *[43, Lemma 2.2] In $R_n \in_U \mathscr{A}$ each vertex is with high probability adjacent to at most $2\frac{\ln(n)}{\ln\ln(n)}$ pendant vertices.*

We define $\mathscr{B}_n = \{G \in \mathscr{A}_n \mid \Delta(G) \geq b\}$, where $b = b(n) = \left\lceil \frac{8}{\alpha} \ln(n) \right\rceil$ and where α is as in Theorem 59 for H being a single vertex. Further, define $\mathscr{B}'_n = \mathscr{B}_n \cap \{G \in \mathscr{A}_n \mid G \text{ has } \geqslant \alpha n \text{ pendant vertices and each vertex is adjacent to } \leqslant 2\frac{\ln(n)}{\ln\ln(n)} \text{ pendant vertices}\}$. We want to show that with high probability $R_n \in_U \mathscr{A}_n$ is not in $\mathscr{B}_n$. We know that with high probability G has at least αn pendant vertices by the Pendant Appearances Theorem (see Theorem 59) and with high probability each vertex is adjacent to at most $2\frac{\ln(n)}{\ln\ln(n)}$ pendant vertices by Lemma 61. It follows that to show that with high probability $R_n \in_U \mathscr{A}_n$ is

not in $\mathscr{B}_n$ it is sufficient to show that with high probability $R_n \in_U \mathscr{A}_n$ is not in $\mathscr{B}'_n$. This is done by constructing many graphs in $\mathscr{A}_n$ from each graph in $\mathscr{B}'_n$ with little double counting, so we can not have started with too many graphs in $\mathscr{B}'_n$.

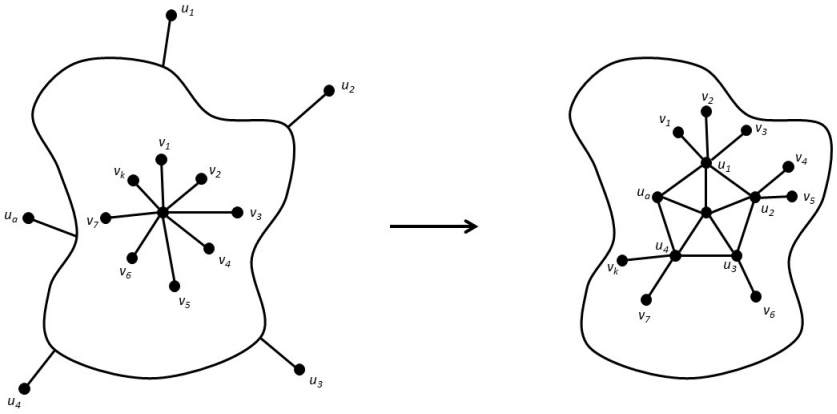


Figure 3.2: $G \in \mathscr{A}_n$ to $G' \in \mathscr{A}_n$

We take a graph $G \in \mathscr{B}'_n$. Then there exists a vertex of degree at least b. We fix an embedding of G in the surface. The construction works as follows: Choose $a = a(n) = \lfloor 2\ln(n) \rfloor$ distinct pendant vertices not adjacent to v, we call them $u_1, \ldots, u_a$. Denote the neighbours of v as $v_1, \ldots, v_k$. Remove the edges incident to v and to $u_1, \ldots, u_a$. Then form a wheel around v with the vertices $u_1, \ldots, u_a$. Then choose a special vertices from $v_1, \ldots, v_k$, say $v_{i_1}, \ldots, v_{i_a}$. For each j, add an edge between u_j and each of $v_{i_j}, \ldots, v_{i_{j+1}-1}$. This finishes the construction of a graph embeddable in the same surface. Since all second neighbours of v in the new graph are the original neighbours of v and each new neighbour of v is a pendant vertex in the original graph, each new graph is not constructed too many times. This then implies that we can not have started with too many vertices in $\mathscr{B}'_n$, as required. ■

With the help of the Pendant Appearances Theorem (see Theorem 59) we can not just deduce results on the maximum degree of a graph chosen uniformly at random from a class with a growth constant, but also bounds on the number of vertices of any chosen constant degree.

Corollary 62. *[45] Let the class $\mathscr{A}$ of graphs have growth constant γ, and let $R_n \in_U \mathscr{A}_n$. Let $k \geq 1$ and suppose that the star with $k-1$ edges, rooted at its centre, can be attached to $\mathscr{A}$. Then there exists a constant $\alpha > 0$ such that*

$$\mathbb{P}[R_n \text{ has } < \alpha n \text{ vertices of degree } k] \leq e^{-\Omega(n)} .$$

Sketch Proof. Let

$$\alpha_k = \frac{d}{\gamma^k \cdot (k+2)!}$$

for $d = \frac{1}{9e^2}$. Then by the proof of the Pendant Appearances Theorem (see Theorem 59), where the graph H is the star on k vertices rooted at the centre vertex, the result follows. ■

From the Pendant Appearances Theorem (see Theorem 59), we can also infer a result on the number of automorphisms a graph chosen uniformly at random from a graph class with a growth constant has.

Corollary 63. *[45] Let the class $\mathscr{A}$ of graphs have growth constant γ, and let $R_n \in_U \mathscr{A}_n$. Then there are constants $a, b > 0$ such that the number of automorphisms* $\mathrm{aut}(R_n)$ *of R_n satisfies*

$$\mathbb{P}(2^{an} \leqslant \mathrm{aut}(R_n)) = 1 - o(2^{-bn}) \ . \tag{3.6}$$

When bounding the size of the graph classes $\mathscr{G}_n^{g(n)}$ and $\mathscr{H}_n^{g(n)}$, we will often try to relate the graphs in this class to planar graphs. For this, we look for smallest planarising sets or non-contractible cycles in the graph that we can "cut" along to simplify the surface. For this reason, we will now state some results on simple cycles in graphs embedded on surfaces and start by defining some properties of simple cycles in graphs embedded in a surface.

Definition. *Let G be a graph that is embedded in an orientable or non-orientable surface* $\mathbf{S}$ *and C a simple cycle in the graph G. If removing the image of C in* $\mathbf{S}$ *separates the surface, then C is called* ***surface-separating*** *in the given embedding. If one of these two parts is homeomorphic to a disc, or equivalently if the embedded C can be continuously deformed to a point, in other words is homotopic to the constant map, then C is said to be* ***contractible*** *in this embedding, otherwise C is called* ***non-contractible****.*

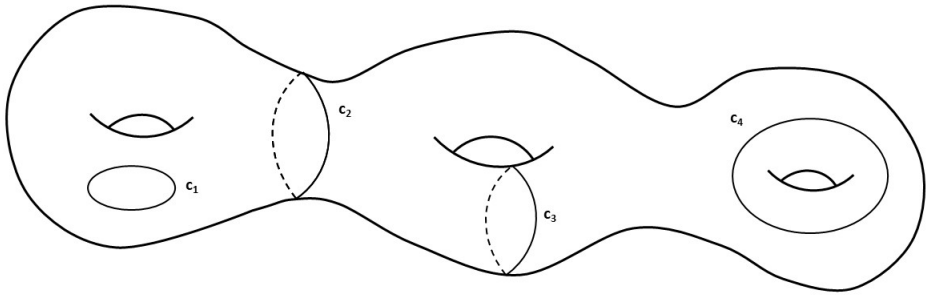


Figure 3.3: c_1 is surface-separating and contractible, c_2 is surface-separating, and c_3 and c_4 are non-surface-separating

Note that any facial walk of a cellularly embedded graph is contractible. Furthermore, by cutting along a surface-separating non-contractible simple cycle and pasting in discs, we can split the original surface into two surfaces whose genera add up to the genus of the original surface. If we cut along a non-surface-separating simple cycle C, the Euler genus of the resulting surface is the Euler genus of the original surface minus two if C is two-sided, and minus one if C is one-sided [48, Lemma 4.2.4]. As a consequence, non-surface-separating, non-contractible simple cycles in graphs embedded on a surface are of great use, as they provide a tool for relating graphs on higher genus surfaces to graphs on lower genus surfaces. To make use of these, we wish to know the maximum length of the shortest non-surface-separating, non-contractible simple cycle in a graph embedded in a surface of genus $g(n)$. This length is called the edge-width of a graph embedding.

Definition. *Let E be a fixed embedding of a graph G in the surface* $\mathbf{S}$*. We define the* ***edge width*** *of E to be the length of the shortest non-contractible simple cycle in E.*

To reduce the genus of a graph by cutting along a curve on the surface it is embedded on, we need not necessarily find a non-contractible simple cycle, but it suffices to find a non-contractible curve in the surface which meets the embedded graph only in vertices. We also give a name to the minimum number of vertices any such curve meets in a fixed embedding.

Definition. *Let E be a fixed embedding of a graph G in the surface* $\mathbf{S}$*. We define the minimum of the number of vertices on a non-contractible curve which meets E only in vertices to be the* ***face-width*** *of the embedding E.*

Note that the name face-width stems from the fact that it is equally the minimum number of faces one has to traverse to close a non-contractible curve. In a similar fashion

to before, cutting along these curves and pasting in discs either reduces the genus of a graph or splits the graph into two parts of smaller genus. Finding tight upper bounds on the maximum face width is also certainly interesting to us in order to find upper bounds on the size of the graph classes $\mathscr{G}_n^{g(n)}$ and $\mathscr{H}_n^{g(n)}$.

3.1.1 Properties of Planar Graphs

In this section, we point out some important basic properties of planar graphs. In the previous section, we have already touched upon some of these properties. Note that a planar graph may have more than one embedding in the plane. We talk about two embeddings of graphs being equivalent if they have the same faces and display the same properties. We define this more rigorously in the following definition.

Definition. *Two embeddings of a connected graph G in the plane are called **equivalent** if for every vertex of the graph in the embedding the cyclic order of the incident edges is the same or each cyclic order is reversed in the two embeddings. Two embeddings of a graph G in the plane are called equivalent if the embeddings of its components are equivalent.*

Intuitively, this means that two embeddings are equivalent if they can be continuously deformed into one another on the sphere or are mirror images of one another. It is worth noting here that embeddings of graphs on the plane and the sphere are equivalent. As a result, for a planar embedding of a graph and for any face f in this embedding, there exists an equivalent embedding in which f is the outer face. We now move on to a special subclass of embeddings, the ones where each face has minimum size, where the size of a face is equal to the number of edges in its facial walk. In a simple graph, the minimum size of a face is at least three.

Definition. *A cellular embedding E of a connected graph G in a surface $\mathbf{S}$ in which every face has size exactly three is called a **triangulation** of $\mathbf{S}$.*

Note that this definition requires a triangulation to be a cellular embedding. Any planar embedding of a graph on $n \geqslant 3$ vertices can be extended to a triangulation in such a way that the resulting graph is still simple. This implies that all maximal, with respect to adding edges, simple planar graphs are triangulations. Since in a triangulation every face is incident to exactly three edges and every edges lies in the boundary of two faces, we know that the maximum number of edges possible in a planar graph occurs when $3f = 2e$, where f is the number of faces in an embedding and e is the number of edges of the graph.

As mentioned before, planar graphs are addable. This means that it is possible to add an edge between any two vertices in different components without destroying the

planarity of the graph; and that a graph is planar if and only if all of its components are planar. By Theorem 58 it would thus suffice to show that the class of planar graphs is small in order to prove that it must have a growth constant.

We have thus far discussed properties that all planar graphs possess and will now move on to give an overview of results on the class of planar graphs and likely properties of a graph chosen uniformly at random from the class of planar graphs. Alain Denise, Marcio Vasconcellos and Dominic J. A. Welsh [28] were the first to study the class of planar graphs. Over recent years, the class of planar graphs has found to be of great interest and has been extensively studied. Today much is known about the size of $\mathscr{P}$ and common properties of a graph chosen uniformly at random from $\mathscr{P}$. Note that when we speak about a random planar graph on the vertex set $[n]$, or equivalently a random n-vertex graph on a surface, we mean a graph chosen uniformly at random from all such graphs on n labelled vertices. In the following, we will summarize some important results on the class of planar graphs. We will make use of some of these results as well as some of the proof ideas later on in the thesis. While most of the precise results on the class of planar graphs are actually obtained using generating function methods, our focus will be on results obtained without the use of generating functions. We have this focus since generating function methods appear not to be easily applicable to graphs embeddable on a surface of non-constant Euler genus.

Most of the results on common properties of a random planar graph make use of bounds on the size of the class of planar graphs. Showing that the class of planar graphs has a growth constant and finding bounds on the size of this constant was hence a major breakthrough in the study of random planar graphs. Colin McDiarmid, Angelika Steger and Dominic JA Welsh [45] were the first to prove the existence of this growth constant, which we call the labeled planar graph growth constant and denote by $\gamma_{\mathscr{P}}$.

Theorem 64. *[45, Theorem 3.2] There exists a finite constant $\gamma_{\mathscr{P}} > 0$ such that*

$$\left(\frac{|\mathscr{P}_n|}{n!}\right)^{\frac{1}{n}} \to \gamma_{\mathscr{P}} \ .$$

Sketch Proof. Note that by Theorem 58 it suffices to show that the class of planar graphs is small to prove the existence of a growth constant, as we have already shown that the class of planar graphs is addable. It was shown in [28] that there exists a constant $0 < c < \infty$ such that the number of unlabelled planar graphs on n vertices is bounded above by c^n. Since the number of labeled planar graphs on n vertices is at most $n!$ times the number of unlabeled planar graphs on n vertices, this result implies that $\mathscr{P}$ is small. Applying Theorem 58 gives the required result. ■

Building on work in [22], Giménez and Noy [34] found in 2008, using a generating function approach, asymptotic estimates for the number of planar graphs on n vertices. Note that generating functions are a very powerful tool for investigating graph classes of graphs embeddable on constant genus surfaces. However, as soon as the genus of the surface is not constant any more, generating function methods appear to break down, which is why we mostly focus on results not involving them in this section. The following theorem is nonetheless stated for completeness.

Theorem 65. *[34] Let $|\mathscr{P}_n|$ be the number of planar graphs on n vertices. Then*

$$|\mathscr{P}_n| \sim c \cdot n^{-\frac{7}{2}} \cdot \gamma_{\mathscr{P}}^n \cdot n! \tag{3.7}$$

where $c \approx 0.4261 \cdot 10^{-5}$ and $\gamma_{\mathscr{P}} \approx 27.2269$.

By now much is known about what a typical random planar graph looks like. Using generating function methods, very precise results about the asymptotic distribution of certain properties of random planar graphs have been obtained. As an example, it is known that the number of pendant vertices, that is vertices of degree one, in a graph chosen uniformly at random from $\mathscr{P}_n$ is asymptotically normally distributed with mean and variance ψn, where $\psi = \gamma_{\mathscr{P}}^{-1} \approx 0.04500$ [34, Theorem 4]. Furthermore, the number of edges in a random planar graph with n vertices is asymptotically normally distributed with mean $\mu_n = \kappa n$ and variance $\sigma_n^2 = \lambda n$ where $\kappa \approx 2.21326$ and $\lambda \approx 0.43034$ [34, Theorem 2]. For the number of components $\kappa(R_n)$ of a graph R_n chosen uniformly at random from all planar graphs on n vertices we also know that $\kappa(R_n) - 1$ is asymptotically Poisson distributed with mean $\nu \approx 0.0374$. Hence, the probability that a random planar graph on n vertices is connected tends to $e^{-\nu} \approx 0.9633$ as n tends to infinity [34, Theorem 6].

These results have all been obtained using generating functions. However, while one can not make as precise statements without the use of generating functions, many results on properties of random planar graphs have been obtained using more combinatorial methods and have been stated in the previous section. Since the class of planar graphs has a growth constant, we know that both the Pendant Appearances Theorem (see Theorem 59) and the bound on the maximum degree in Theorem 60 hold for random planar graphs.

While the asymptotic distribution of the degrees is known by using generating function methods, one can also obtain results about the number of vertices of any given fixed degree using more fundamental methods. Corollary 62 states that for every fixed constant $k \geqslant 2$, there are, with exponentially small failure probability in the limit, a linear number of vertices of degree $k - 1$.

3.1.2 Properties of Graphs on Surfaces

A natural extension of the class of planar graphs is the class of graphs embeddable in any fixed surface $\mathbf{S}$. Colin McDiarmid [42] was the first to show that the class of graphs embeddable in any fixed surface, $\mathscr{G}_n^h$, has the same growth constant as planar graphs, that is for any fixed constant h and as $n \to \infty$

$$\left(\frac{\left|\mathscr{G}_n^h\right|}{n!}\right)^{\frac{1}{n}} \to \gamma_{\mathscr{P}} \ .$$

By the same argument it can also be shown that

$$\left(\frac{\left|\mathscr{H}_n^h\right|}{n!}\right)^{\frac{1}{n}} \to \gamma_{\mathscr{P}} \ .$$

For graphs on surfaces of constant genus, Chapuy et al [26] later managed to use a generating function approach to find asymptotic estimates for the size of $\mathscr{G}_n^h$ and showed that for any Euler genus $h \in 2\mathbb{N}$ in the orientable case or $h \in \mathbb{N}$ in the non-orientable case the sizes of the classes of graphs of orientable or non-orientable Euler genus at most h satisfy the following

$$\begin{aligned} \left|\mathscr{G}_n^h\right| &\sim c_h \cdot n^{\frac{5(h-2)}{4}-1} \cdot \gamma_{\mathscr{P}}^n \cdot n! \\ \left|\mathscr{H}_n^h\right| &\sim c'_h \cdot n^{\frac{5(h-2)}{4}-1} \cdot \gamma_{\mathscr{P}}^n \cdot n! \ , \end{aligned} \tag{3.8}$$

where c_h and c'_h are positive constants depending only on the Euler genus h of the surface.

Some of the properties that planar graphs possess carry over to the more general setting of graphs on surfaces, but not all. While we can still extend any simple graph embedded on a surface to a triangulation by adding edges, it may now not be possible to do so in a way that the resulting triangulation is simple. As a result, unlike in the planar case, edge-maximal simple graphs on surfaces are not necessarily triangulations. As an example, the graph $K_8 - E(C_5)$ is an edge-maximal toroidal graph but is not a toroidal triangulation [36]. It can however be shown that any edge maximal graph on a surface is not too far from being a triangulation. Recently, Colin McDiarmid and David R. Wood [47] proved that for every fixed surface Σ of Euler genus g, every edge-maximal embedding of a graph is at most $O(g)$ edges short of a triangulation of Σ.

Similarly, decomposability does not carry over from planar graphs to graphs on surfaces. While K_5, the complete graph on five vertices, is embeddable in the torus, two copies of K_5 are not embeddable in a single torus. However, graphs on surfaces are still bridge-addable, even when the genus of the surface grows with the number of vertices.

In the following, we will usually treat the orientable and non-orientable cases separately. The following observation gives a straightforward relationship between the size

of the class of graphs of bounded non-orientable genus and the size of class of graphs of bounded orientable genus.

Observation 66. *A graph embeddable in an orientable surface* $\mathbf{S}$ *of Euler genus* g *can be embedded in a non-orientable surface of Euler genus* $g+1$ *(by adding a crosscap to the surface* $\mathbf{S}$*).*

An example of the case where the non-orientable Euler genus is exactly one larger than the orientable Euler genus is the complete graph on seven vertices, K_7. It has orientable Euler genus equal to two and non-orientable Euler genus equal to three. Note however, that no such reverse implication is possible. Graphs of non-orientable genus at most g can have arbitrarily large orientable genus. This was shown by J.R. Fiedler et al [31], who proved that graphs of non-orientable genus one can have arbitrarily large orientable genus.

It turns out that asymptotically for large n, a random graph on a fixed surface looks the same as a random planar graph. Note that the Pendant Appearances Theorem (see Theorem 59), giving that with exponentially small failure probability a random graph from the class has linearly many pendant appearances of a fixed planar graph, and Theorem 60, showing that the maximum degree of a random graph is with high probability of order $O(\log(n))$, still hold for graphs embeddable in a fixed surface. Similarly, the results on the number of components in a random graph also hold for the class of graphs embeddable in any fixed surface.

Furthermore, asymptotic distribution results found using generating functions are also the same as in the planar case. In [26], it was shown that the number of edges in a random graph on a fixed surface is asymptotically normal with mean and variance exactly the same as in the planar case. Similarly, the number of components and the number of pendant appearances of a fixed graph also have the same asymptotic distribution as in the planar case. Therefore, random graphs on a fixed surface behave in many respects asymptotically the same as planar graphs.

A natural question arising from this is to ask what happens when the surface is not kept constant, but the genus of the surface grows with the number of vertices. Equivalently, what happens when we look at the class of graphs embeddable in the surface of Euler genus $g(n)$, where n is the number of vertices of the graph and where $g(n)$ is some non-negative, non-decreasing function of n. In this case, analytic methods currently seem to fall down and more fundamental methods such as double counting need to be used. The case when $g(n)$ is a non-constant function depending only on the number of vertices

is studied in this chapter of the thesis. We show that for any given $g(n) = o\left(\frac{n}{\log^3(n)}\right)$, the classes of graphs $\mathscr{G}_n^{g(n)}$ and $\mathscr{H}_n^{g(n)}$ have growth constants equal to the planar graph growth constant $\gamma_{\mathscr{P}}$. Additionally, the classes of graphs embeddable in an orientable or non-orientable surface of Euler genus at most any given $g(n) = \Theta\left(\frac{n}{\log(n)}\right)$ is small, while the classes of graphs $\mathscr{G}_n^{g(n)}$ and $\mathscr{H}_n^{g(n)}$ for any given $g(n) = \omega\left(\frac{n}{\log(n)}\right)$ are over-large. Finally, we investigate some common properties of graphs chosen uniformly at random from these classes and give an outlook on further open problems concerning graph classes on non-constant surfaces.

3.2 Basic Results on Graphs on Surfaces

In this section, we collect some useful theorems and observations on graphs on surfaces, which we will repeatedly make use of in the remainder of this thesis. One crucial formula is Euler's formula for general surfaces.

Theorem 67. ***(Euler's formula)*** *Let E be a cellular embedding of a finite, connected graph G on a surface of Euler genus g. Denote by v the number of vertices, by e the number of edges and f the number of faces of E. Then E satisfies*

$$2 - g = v - e + f \ .$$

Note that this formula also holds for cellularly embedded pseudographs, that is graphs with multi-edges or self-loops. We now extend this theorem to disconnected graphs. To do so, we need to define a similar notion to being cellularly embedded for disconnected graphs.

Definition. *Let G be a graph with $\kappa = \kappa(G) \geq 1$ components $G_1, \ldots, G_\kappa$. A **disconnected cellular embedding** γ of G is specified by a family of embeddings $\gamma_1, \ldots \gamma_\kappa$, where γ_i is a cellular embedding of G_i, each with a distinguished 'outer' face and where all faces in the embedding apart from the 'outer' face are 2-cells. We think of the outer faces as forming a single shared outer face (which is not a 2-cell if $\kappa \geq 2$, with boundary in κ parts). Further, the Euler genus g of the surface that γ is an embedding on satisfies $g = \sum_i g_i$, where g_i is the Euler genus of γ_i.*

We may justify this definition as follows: For each i let x_i be an arbitrary vertex on the outer face of G_i, and let T be an arbitrary tree on $x_1, \ldots, x_\kappa$. If we add the edges of T in the outer face of the embedding in a natural way then the κ parts of the boundary merge into one facial walk, and we obtain a cellular embedding γ' of the connected graph $G' = G \cup E(T)$. We have that the new number of vertices v' satisfies $v' = v$, the new

number of edges e' satisfies $e' = e + \kappa - 1$ and the new number of faces f' satisfies $f' = \sum_i (f_i - 1) + 1 = \sum_i f_i - (\kappa - 1)$. The Euler genus of γ' is thus equal to

$$\begin{aligned} 2 + e' - v' - f' &= 2 + \left(\sum_i e_i + (\kappa - 1)\right) - \left(\sum_i v_i\right) - \left(\sum_i f_i + (\kappa - 1)\right) \\ &= \sum_i (e_i - v_i - f_i + 2) \\ &= \sum_i g_i \ . \end{aligned}$$

It follows that, however we choose the outer face, the vertices x_i and the tree T, the embedding γ' has Euler genus $\sum_i g_i$. By Euler's formula for connected graphs (see Theorem 67), each component of G satisfies

$$v_i - e_i + f_i = 2 - g_i \ .$$

Regarding the outer faces as shared, the number f of faces of γ is

$$\sum_i (f_i - 1) + 1 = \sum_i f_i - (\kappa - 1) \ .$$

So for γ we have

$$\begin{aligned} v - e + f &= \sum_i (v_i - e_i + f_i) - (\kappa - 1) \\ &= \sum_i (2 - g_i) - (\kappa - 1) \\ &= \kappa + 1 - g \end{aligned}$$

and so

$$\begin{aligned} v - e + f - \kappa(G) &= 1 - \sum_i g_i \\ &= 1 - g \ . \end{aligned}$$

This is Euler's equation generalised to graphs which need not be connected. We state this result formally in the following theorem.

Theorem 68. ***(Euler's formula for disconnected graphs)*** *Let γ be a disconnected cellular embedding of a graph G with κ components on a surface of orientable Euler genus g. Denote by v the number of vertices of G, by e the number of edges of G and by f the number of faces of γ. Then*

$$1 - g = v - e - \kappa + f \ .$$

A similar statement holds in the non-orientable case.

From Euler's formula, we can deduce many other results that hold for graphs embedded in a surface. By using the fact that for all embeddings of simple graphs $3f \leqslant 2e$, since each face has at least three edges and each edge appears twice on a face, and combining it with Euler's formula (see Theorem 67) we get the following result.

Observation 69. *The maximum number of edges a simple planar graph on $n \geqslant 3$ vertices can have is $3n - 6$.*

Note further that all edge maximal planar graphs on n vertices have exactly this number of edges, since it is always possible to extend a simple planar graph to a simple triangulation of the plane. Similarly to the planar case, no simple graph embeddable on a surface can have more edges than a triangulation of said surface, as a result it always holds that $3f \leqslant 2e$. Combined with Euler's formula (see Theorem 67), this again gives a bound on the maximum number of edges a graph embedded on a surface can have. Note however that unlike in the planar case, not all edge-maximal graphs embeddable on this surface have exactly this many edges.

Observation 70. *Each simple graph on n vertices embeddable on a surface of Euler genus h has at most $3(n + h - 2)$ edges.*

We now know the maximum number of edges an n-vertex graph of genus at most g can have. Euler's formula (see Theorem 67) can also be used to help us find the minimum number of edges a connected graph on n vertices needs to have to have genus at least g. Note that any cellular embedding of a graph (with at least one edge) on a surface of Euler genus g has at least one face, that is $f \geqslant 1$. Furthermore, note that any minimum genus embedding of a connected graph must necessarily be cellular as it would otherwise be possible to cut along the boundary cycle of the non-cellular face and replace it with a disc.

Definition. *Given a graph G, a **minimum Euler genus orientable embedding** is an embedding of G on an orientable surface such that there exists no embedding of G on an orientable surface of smaller Euler genus. We define the **minimum orientable Euler genus** to be the orientable Euler genus of said surface. Similarly, a **minimum Euler genus non-orientable embedding** in an embedding of G on a non-orientable surface such that there exists no embedding of G on a non-orientable surface of smaller Euler genus (unless G is planar, in which case any planar embedding of the graph is also a non-orientable minimum genus embedding). As before, the **minimum non-orientable Euler genus** is defined to be the non-orientable Euler genus of this surface. The **minimum Euler genus** of a graph is defined to be the minimum of the minimum orientable and minimum non-orientable Euler genus.*

Observation 71. *Let $h \geq 2$ be even. If a connected graph G is in $\mathscr{G}_n^h$ and not in $\mathscr{G}_n^{h-2}$ then each embedding of G in $\mathbf{S}_{h/2}$ is cellular. Similarly, let $h \geqslant 1$. If a connected graph G is in $\mathscr{H}_n^h$ but not in $\mathscr{H}_n^{h-1} \cup \mathscr{G}_n^{h-1}$ then each embedding of G in $\mathbf{N}_h$ is cellular.*

Using this observation and Euler's formula (see Theorem 67) on connected graphs we obtain the following result.

Observation 72. *Suppose G is a connected graph on n vertices and of Euler genus at least g. Then G has at least $g + n - 1$ edges.*

As mentioned before, bounds on the maximum edge and face width of a graph embedded in a surface of Euler genus g are of importance to us when trying to relate graphs on surfaces to planar graphs. We will now state some known results on the edge and face width of graphs which we will repeatedly make use of later on.

Theorem 73. *[27] There exist constants c and c' such that the following holds. Let E be a triangulation of a not necessarily simple, connected n-vertex graph G on a surface of Euler genus $g \geqslant 1$. Then in E there exists some non-contractible simple cycle in G of length at most*

$$c \cdot \sqrt{\frac{n}{g}} \cdot \log(g)$$

and some non-separating simple cycle in G of length at most

$$c' \cdot \sqrt{\frac{n}{g}} \cdot \log(g) \ .$$

Furthermore, these bounds are tight up to a constant.

While the second statement implies the first when $c = c'$, in [27] the constant c is strictly smaller than the constant c'. Theorem 73 readily implies that the face-width of any n-vertex graph cellularly embedded in a surface of Euler genus g is at most $c \cdot \sqrt{\frac{n}{g}} \cdot \log(g)$. Similarly, since any graph on a surface can be extended to a, not necessarily simple, triangulation, for any n-vertex graph embedded in a surface of Euler genus g, there exists a non-separating curve on the surface meeting the graph in at most $c' \cdot \sqrt{\frac{n}{g}} \cdot \log(g)$ vertices and nowhere else.

We have just discussed one way of reducing the genus of a graph or a surface, which is by cutting the surface along a non-separating or non-contractible curve. Another way of reducing the genus of a graph is to find bounds on the minimum size of a set of edges or vertices we need to delete in order to reduce the genus of said graph. If we instead look at which vertices or edges we need to delete in order to obtain a planar graph, this is called a planarising vertex or planarising edge set respectively. It is certainly not obvious

what the best set of vertices or edges to delete is in order to reduce the genus of a graph. As a result, there has been great interest in algorithms for finding such small planarising sets efficiently. Currently no efficient algorithm for finding the smallest vertex or edge planarising set is known and the problems have been shown to be NP-complete [58]. In our case, however, it suffices to find an upper bound on the size of such planarising sets, rather than finding the explicit set. We will repeatedly make use of the following bounds on the maximum size of vertex and edge planarising sets.

Theorem 74. *[30, Theorem 2, Theorem 6] Let G be an n-vertex graph of Euler genus g. Then there exists a set of no more than $2\sqrt{2gn}$ vertices of G whose removal leaves a planar graph.*

Theorem 75. *[30, Theorem 4] Let G be a connected n-vertex graph of maximum degree d and Euler genus g. There exists a set of no more than $4\sqrt{dg(n+g-2)}$ edges whose removal makes G planar.*

Note that the order of these bounds is certainly tight. To show this, we first need to make a definition.

Definition. *A balanced vertex separating set of a graph on n vertices is a set of vertices whose removal leaves no component with more than $2n/3$ vertices. Similarly a balanced edge separating set of a graph on n vertices is a set of edges whose removal leaves no component with more than $2n/3$ vertices.*

We can now prove the tightness of the results in Theorems 74 and 75. In the orientable case, this can be seen by the following example. For graphs G on n vertices, m edges, genus g and of maximum degree d, there are examples with the smallest balanced vertex separating set of size $\Omega\left(\sqrt{gn}\right)$ by [33] and the smallest balanced edge separating set of size $\Omega\left(\sqrt{dgm}\right)$ by [53]. Suppose it was possible to planarise G by removing $o\left(\sqrt{gn}\right)$ vertices or $o\left(\sqrt{dgm}\right)$ edges. Then we could find in the resulting planar graph a balanced vertex separating set of size $O\left(\sqrt{n}\right)$ by [38] and a balanced edge separating set of size $O\left(\sqrt{dn}\right)$ by [29]. This implies that, assuming $g(n) \to \infty$, it would be possible to find in G a balanced vertex separating set of size $o\left(\sqrt{gn}\right)$ and a balanced edge separating set of size $o\left(\sqrt{dgm}\right)$, contradicting the above lower bounds on the size of separating sets in the case where g is non-constant. The order of the bounds in Theorem 74 and Theorem 75 are also tight in the non-orientable case, which simply follows from the orientable case using Observation 66. This tightness result implies that for genus $g(n) \geq n$, the smallest planarising vertex set has size $\Omega(n)$. For large genus, the smallest planarising vertex set is hence too large to help in trying to compare the size of the large genus graph classes to planar graphs. However, the smallest edge planarising set will turn out to be useful

when trying to bound the size of $\mathscr{G}_n^{g(n)}$ above for some $g(n)$ by relating it to planar graphs.

We now further make some observations about the growth of the graph classes $\mathscr{G}_n^{g(n)}$ and $\mathscr{H}_n^{g(n)}$ when going from n to $n+1$ vertices, in the different cases when the genus also increases or does not increase when adding a vertex (see also Lemma 116).

Lemma 76. *Note that for any $k \geqslant 0$ and $n \in \mathbb{N}$ we have for $\mathscr{A}$, where $\mathscr{A}^k = \mathscr{G}^k$ or $\mathscr{A}^k = \mathscr{H}^k$, that*

$$\begin{aligned} \left|\mathscr{A}_{n+1}^k\right| &\geqslant (n+1) \cdot \left|\mathscr{A}_n^k\right| \quad \textit{and} \\ \left|\mathscr{A}_{n+1}^{k+2}\right| &\geqslant \frac{n^2}{2} \cdot \left|\mathscr{A}_n^k\right| \ . \end{aligned} \tag{3.9}$$

Proof. We prove the two bounds by a simple double counting argument. For the first case, suppose that G is a graph in $\mathscr{A}_n^k$.

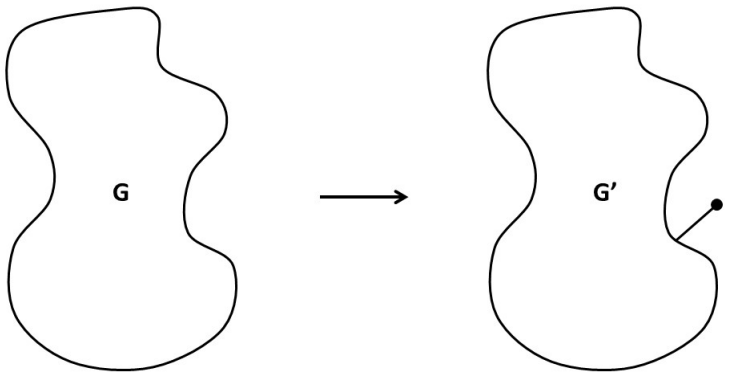


Figure 3.4: $G \in \mathscr{A}_n^k$ to $G' \in \mathscr{A}_{n+1}^k$

By adding the vertex with label $n+1$ to G as a pendant vertex adjacent to any of the present vertices, or as an isolated vertex, we construct a graph G' contained in $\mathscr{A}_{n+1}^k$. We have n choices of where to attach the vertex $n+1$ to as a pendant vertex and one choice for adding vertex $n+1$ as an isolated vertex, so we can construct at least $n+1$ graphs G' from G.

How often is each graph G' constructed? By deleting the vertex $n+1$ from G' we return uniquely to G, that is each graph G' is constructed only once and it follows that

$$\left|\mathscr{A}_{n+1}^k\right| \geqslant (n+1) \cdot \left|\mathscr{A}_n^k\right| \ , \tag{3.10}$$

as required.

To prove the second inequality, we again take a graph $G \in \mathscr{A}_n^k$ and attach a new vertex with label $n+1$ to the graph.

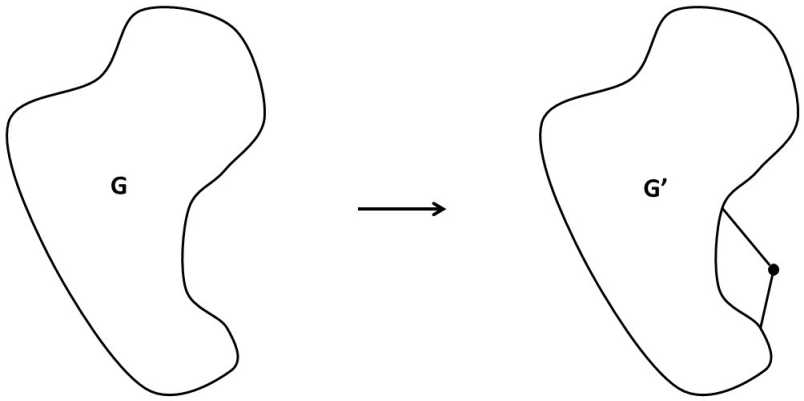


Figure 3.5: $G \in \mathscr{A}_n^k$ to $G' \in \mathscr{A}_{n+1}^{k+1}$

This time $n+1$ is either attached to any two of the previous vertices, or as a pendant vertex, or as an isolated vertex. Note that adding two edges between the vertex $n+1$ and any other vertex increases the Euler genus of the graph by at most two, since we could simply attach a handle to the surface in two faces incident to the two vertices we wish to connect vertex $n+1$ to and then add vertex $n+1$ along this handle. The resulting graph, call it G', is thus certainly contained in $\mathscr{A}_{n+1}^{k+2}$. By this method, we construct from each graph G exactly $\binom{n}{2} + n + 1 > \frac{n^2}{2}$ graphs G'.

How often is each graph G' constructed? By deleting vertex $n+1$ from G' we uniquely recover G. Each graph G' is thus constructed exactly once and it follows that

$$\left|\mathscr{A}_{n+1}^{k+2}\right| \geqslant \frac{n^2}{2} \cdot \left|\mathscr{A}_n^k\right| \ , \tag{3.11}$$

as required. ■

We can find a similar bound in the case where we just add a handle to the surface.

Observation 77. *For any $k(n) \geqslant 0$ such that $k(n) = o(n^2)$ we have that*

$$\left|\mathscr{G}_n^{k(n)+2}\right| \geqslant \frac{n^2}{6(k(n)+n)} \cdot \left|\mathscr{G}_n^{k(n)}\right| \cdot (1+o(1)) \text{ as } n \to \infty \tag{3.12}$$

and

$$\left|\mathscr{H}_n^{k(n)+2}\right| \geqslant \frac{n^2}{6(k(n)+n)} \cdot \max\left\{\left|\mathscr{G}_n^{k(n)}\right|, \left|\mathscr{H}_n^{k(n)}\right|\right\} \cdot (1+o(1)) \text{ as } n \to \infty \ . \tag{3.13}$$

Proof. We again prove this bound by a simple double counting argument. By adding a handle or twisted handle to the surface, we can add any of the non-edges as an edge to the graph. This is since we can attach the handle to the surface inside two faces incident to the two vertices we wish to connect and then add the edge between these vertices along the handle or twisted handle.

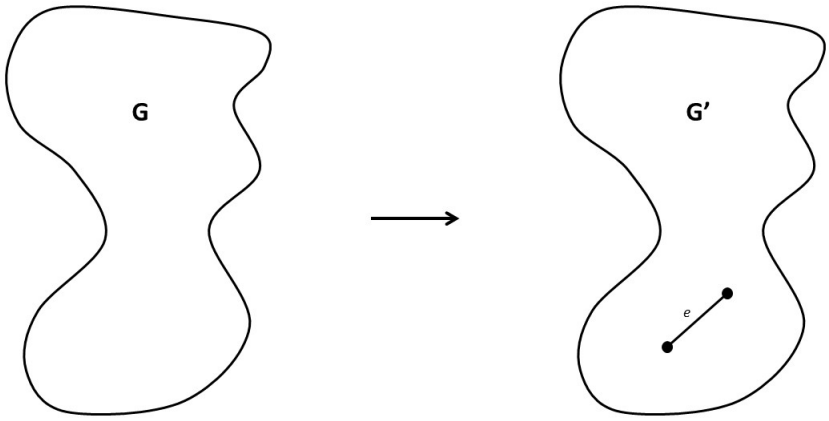


Figure 3.6: $G \in \mathscr{A}_n^{k(n)}$ to $G' \in \mathscr{A}_n^{k(n)+2}$

Since we are restricting ourselves to simple graphs, we are only allowed to add edges that are not already present in the graph. There are at least $\binom{n}{2} - 3n - 3k(n)$ non-edges in the graph, since each graph embedded, not necessarily cellularly, in a surface of genus $k(n)$ has by Euler's formula (see Theorem 67) at most $3n + 3k(n) - 6$ edges. This means that from each graph $G \in \mathscr{A}_n^{k(n)}$ we construct at least $\binom{n}{2} - 3n - 3k(n)$ graphs $G' \in \mathscr{A}_n^{k(n)+2}$. Furthermore, each graph G' has at most $3n + 3k(n)$ edges and as a result is constructed at most this many times. As a result, we get the following inequality

$$\left|\mathscr{A}_n^{k(n)+2}\right| \geqslant \frac{\binom{n}{2} - 3n - 3k(n)}{3n + 3k(n)} \cdot \left|\mathscr{A}_n^{k(n)}\right| \ , \tag{3.14}$$

from which the required bound follows. ■

Finally, we show that in the orientable case a graph chosen uniformly at random from $\mathscr{G}_n^{g(n)}$, as long as $g(n) = o(n^2)$, has with high probability a minimum genus embedding in $\mathbf{S}_{g(n)/2}$ (or $\mathbf{S}_{(g(n)-1)/2}$ if $g(n)$ is odd), which is equivalent to having a cellular embedding if G is connected and a disconnected cellular embedding if G is not connected. Similarly, we show that in the non-orientable case a graph chosen uniformly at random from $\mathscr{H}_n^{g(n)}$, as long as $g(n) = o(n^2)$, has with high probability a minimum genus embedding in either $\mathbf{N}_{g(n)}$ or $\mathbf{N}_{g(n)-1}$, which is equivalent to having a cellular embedding if G is connected and a disconnected cellular embedding if G is disconnected.

Lemma 78. *A graph R_n chosen uniformly at random from the class $\mathscr{G}_n^{g(n)}$ for $g(n) = o(n^2)$ is with high probability not in $\mathscr{G}_n^{g(n)-2}$. Hence, R_n has with high probability a cellular or disconnected cellular embedding in $\mathbf{S}_{g(n)/2}$ or $\mathbf{S}_{(g(n)-1)/2}$.*

Proof. We prove this lemma by a double counting argument. For this define

$$\mathscr{A}_n = \mathscr{G}_n^{g(n)}$$
$$\mathscr{B}_n = \mathscr{G}_n^{g(n)-2} .$$

To prove the lemma it suffices to show that $|\mathscr{B}_n| / |\mathscr{A}_n| \to 0$ as n tends to infinity. This follows directly from applying Observation 77 with $k(n) = g(n) - 2$ and using the fact that

$$\frac{n^2}{6(k(n)+n)} \to \infty$$

as $n \to \infty$ since $g(n) = o(n^2)$.

■

A similar statement can be made in the non-orientable case, again noting that a minimum genus embedding of a connected graph is necessarily a cellular embedding and a minimum genus embedding of a disconnected graph is necessarily a disconnected cellular embedding.

Lemma 79. *A graph chosen uniformly at random from the class $\mathscr{H}_n^{g(n)}$ for $g(n) = o(n^2)$ is with high probability not in $\mathscr{G}_n^{g(n)-2} \cup \mathscr{H}_n^{g(n)-2}$ and so has Euler genus $g(n)$ or $g(n)-1$.*

Proof. This proof is done exactly the same way as in the orientable case, in other words by noting that adding an edge to a graph of non-orientable Euler genus at most $g(n)-2$ increases the Euler genus by at most two. We define

$$\mathscr{A}_n = \mathscr{H}_n^{g(n)}$$
$$\mathscr{B}_n = \{G \in \mathscr{A}_n \mid G \text{ has Euler genus at most } g(n)-2\} = \mathscr{H}_n^{g(n)-2} \cup \mathscr{G}_n^{g(n)-2}$$

and the rest of the proof is exactly the same as in the orientable case, that is the proof of Lemma 78. ■

In a later section, we will construct graphs embeddable in a surface $\mathbf{S}$ by adding edges in the unique face of unicellular maps on $\mathbf{S}$. We wish to show that all connected graphs cellularly embeddable in this surface can be constructed in this way. To do so we need to show that it is possible to get to a unicellular embedding from a cellularly embedded graph. This is where the following two lemmas will come in useful.

Lemma 80. *Suppose G is a graph cellularly embedded in a surface* $\mathbf{S}$ *with more than one face in the embedding. Then there exists an edge e in G such that e is incident to exactly two distinct faces in the embedding.*

Proof. Suppose G is a graph cellularly embedded in a surface $\mathbf{S}$ and fix this embedding. Suppose further that the embedding of G has more than one face and no edge of G is incident to more than one face. Then, since the surface $\mathbf{S}$ is connected, there exists a vertex v incident to at least two faces, F_1 and F_2 say. We now look at the local area around the vertex v, which is locally homeomorphic to a disc. This means that as long as we choose a sufficiently small disc around v, the graph looks like a wheel as is shown in Figure 3.2.

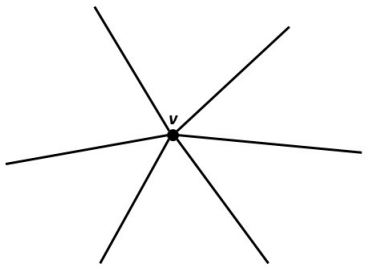


Figure 3.7: Wheel around vertex v

Note that the segment between any two of the edges attached to v must belong to a face, so one of them must belong to F_1 and another to F_2. An example of this is shown in Figure 3.2.

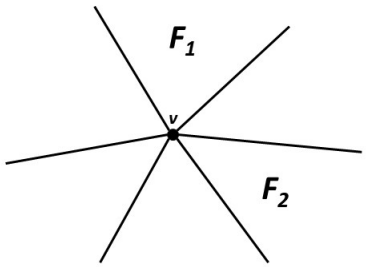


Figure 3.8: F_1 and F_2 incident to v

If we now move anticlockwise starting from F_1 to the next segment, this segment must also belong to F_1, since we have assumed that there is no edge adjacent to two disjoint faces.

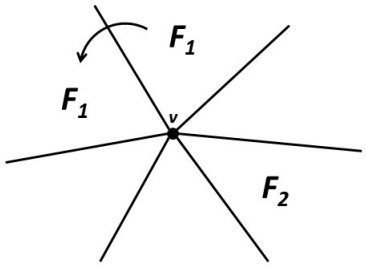


Figure 3.9: Moving anticlockwise starting at F_1

However, continuing in this fashion, we must eventually arrive at a segment belonging to F_1 with the following segment in the anticlockwise ordering belonging to F_2 (as we have assumed that there is a segment belonging to F_2 incident to v). However, this implies that the edge between these two segments is either adjacent to two distinct faces or $F_1 = F_2$, both leading to a contradiction. Hence, any graph embedded in a surface with at least two faces in the embedding must contain an edge incident to two disjoint faces in this embedding. ■

We now know that in each embedding with more than one face there is an edge incident to two disjoint faces. We further need to know for our construction later on that deleting this edge will result in a cellular embedding which has one face less than the original embedded graph. To prove this, we will make use of rotation systems. Since we will not make use of these again in the remainder of the thesis, we will only give a brief definition here. For a more in depth introduction and properties of rotation systems, see Chapter 3 of [48].

Definition. *[48] Assume that G is cellularly embedded in $\mathbf{S}$. Let $\pi = \{\pi_v \mid v \in V(G)\}$ where π_v is the cyclic permutation of the edges incident with the vertex v such that $\pi_v(e)$ is the successor of e in the clockwise ordering around v. The cyclic permutation π_v is called the **local rotation** at v, and the set π is the **rotation system** of the given embedding of G in $\mathbf{S}$.*

Note that every cellular embedding of a graph G in an orientable surface is uniquely determined, up to homeomorphism, by its rotation system. We now extend the notion of the rotation system in order to include embeddings in non-orientable surfaces in the following definition.

Definition. *[48] Let G be a connected graph cellularly embedded in a surface $\mathbf{S}$. Define C_v for each $v \in V(G)$ be an open disc containing v but intersecting only edges incident with v. Further define C_e for each $e \in E(G)$ to be a simple closed polygonal curve such that e is in $\text{int}(C_e)$ and $\text{int}(C_e) \cap \text{int}(C_f) = \emptyset$ if e, f are nonadjacent edges, and $\text{int}(C_e) \cap \text{int}(C_f) = \text{int}(C_v)$ if both e and f are incident with the vertex v. For each edge $e = uv$ of G we check if the clockwise ordering at v and u agree in the disc $\text{int}(C_e)$. If this is the case, then we set $\lambda(e) = 1$, otherwise $\lambda(e) = -1$. This defines a mapping $\lambda : E(G) \to \{+1, -1\}$, called a **signature**. The pair π, λ, where $\pi = \{\pi_v \mid v \in V(G)\}$ is the set of chosen local rotations, is an **embedding scheme** for the given embedding of G.*

Unlike in the orientable case however, the embedding scheme corresponding to an embedding of G is not uniquely determined. We can now prove our next Lemma.

Lemma 81. *Considering a cellularly embedded connected graph G. When deleting an edge e which is incident to two disjoint cellular faces f_1 and f_2 the resulting graph G' is still connected, the induced embedding is still cellular and has one face less than the embedding of G.*

Proof. Suppose we have a cellular embedding of the connected simple graph $G = (V, E)$ in the surface $\mathbf{S}$, with $f \geqslant 2$ faces. Let $e = xy$ be an edge in two distinct faces with facial walks P and Q. From the previous lemma, Lemma 80, we know that such an edge exists. Let $G' = (V, E')$ be the graph obtained from G by deleting e. Then the natural inherited cellular embedding of G' has $f - 1$ faces, namely the faces of G except that the faces with facial walks P and Q have been replaced by a new face with facial walk R. To show this, we first consider the case when $\mathbf{S}$ is orientable.

When $\mathbf{S}$ is orientable, let $\pi = \{\pi(v) : v \in V\}$ be the rotation system for the embedding. The facial walks P and Q must pass through e in different directions, say xy in P and yx in Q. Let P contain the edge-sequence ax, xy, yb (consecutively) and let Q contain cy, yx, xd. Then in $\pi(x)$ we have ax, yx, dx (consecutively) and in $\pi(y)$ we have cy, xy, by. Let $\pi' = \{\pi'(v) : v \in V\}$ be the rotation system for G' obtained by deleting xy and yx from $\pi(x)$ and $\pi(y)$. Thus in $\pi'(x)$ we have ax, dx and in $\pi'(y)$ we have cy, by. Then in G', π' we have the facial walk R constructed as follows: Follow P from the start

to ax then follow Q from xd to cy, then follow P from yb to the end. Other facial walks are as for G, π.

Now suppose $\mathbf{S}$ is non-orientable. Let π, λ form an embedding system for the embedding of G, where $\pi = \{\pi(v) : v \in V\}$ and $\lambda = \{\lambda(e) : e \in E\}$. By reversing a facial walk if necessary, we may assume that the facial walks P and Q pass through e in different directions, say xy in P and yx in Q. The mode of P as it starts e must be the same as the mode of Q as it leaves e (since otherwise reversing Q gives P, contradicting distinctiveness); and thus also the mode of P as it leaves e must be the same as the mode of Q as it starts e.

Let P contain ax, xy, yb and let Q contain cy, yx, xd. Suppose that the mode of P as it starts e is $+1$ (clockwise) and so the mode of Q as it leaves e is $+1$ (note that the other case is very similar, so we will not spell it out here). Suppose also that $\lambda(e) = +1$ (we will spell out the other case later).

In $\pi(x)$ we have ax, yx, dx and in $\pi(y)$ we have cy, xy, by. Let $\pi' = \{\pi'(v) : v \in V\}$ be the rotation system for G' obtained by deleting xy (and yx) from $\pi(x)$ and $\pi(y)$. Then in $\pi'(x)$ we have ax, dx and in $\pi'(y)$ we have cy, by. Let $\lambda' = \{\lambda(e) : e \in E'\}$. Then in G', π', λ' we have the facial walk R constructed as follows: Follow P from the start to ax, then follow Q from xd to cy, then follow P from yb to the end. This works as at x since P is in mode $+1$ as it reaches x and Q is in mode $+1$ as it leaves x; and it works at y since Q is in mode $+1$ as it reaches y and P is in mode $+1$ as it leaves y. Other facial walks are as for G, π, λ.

Now suppose that $\lambda(e) = -1$. Let π' and λ' be formed as before. In $\pi(x)$ we have ax, yx, dx as before, but in $\pi(y)$ we have by, xy, cy. In G', π', λ' we have the facial walk R constructed as follows: Follow P from the start to ax, then follow Q from xd to cy, then follow P from yb to the end. This works at x since P is in mode $+1$ as it reaches x and Q is in mode $+1$ as it leaves x; and it works at y since Q is in mode (-1) as it reaches y and P is in mode (-1) as it leaves y. Again, other facial walks are as for G, π, λ. ■

In the following, we will repeatedly make use of the observations and results presented in this section. We now move on to finding upper and lower bounds on the size of the graph classes $\mathscr{G}_n^{g(n)}$ and $\mathscr{H}_n^{g(n)}$.

3.3 The Size of Graph Classes

This section focuses on finding upper and lower bounds on the size of the graph classes $\mathscr{G}_n^{g(n)}$ and $\mathscr{H}_n^{g(n)}$ for any non-negative function $g(n)$. Before starting with our considerations, we show that it is sufficient to consider only genus functions that are sufficiently

well behaved. Note that when we talk about genus, we will usually refer to the Euler genus rather than the orientable or non-orientable genus unless otherwise specified. For any real-valued function $g = g(n)$ defined on the non-negative integers, define the function $\hat{g}$ by setting

$$\hat{g}(n) = \max\left\{0, \max\{g(m) : 1 \leqslant m \leqslant n\}\right\} .$$

Clearly then $g(n) \leqslant \hat{g}(n)$, and the function $\hat{g}(n)$ is non-negative and non-decreasing.

Lemma 82. *Let $h = h(n)$ be non-decreasing and satisfy $h(n) \to \infty$ as $n \to \infty$. If $g(n) = o(h(n))$ then $\hat{g}(n) = o(h(n))$, and if $g(n) = O(h(n))$ then $\hat{g}(n) = O(h(n))$.*

Proof. Suppose that $g(n) = o(h(n))$ and let $\epsilon > 0$. Then there exists $n_0 \in \mathbb{N}$ such that $g(n) \leqslant \epsilon h(n)$ for each $n \geqslant n_0$. Now let $M = \max\{0, g(1), \ldots, g(n_0 - 1)\}$. Then $\hat{g}(n) \leqslant \max\{M, \epsilon h(n)\}$ for each n. Since $h(n) \to \infty$ as $n \to \infty$, we can choose a $n_1 \in \mathbb{N}$ sufficiently large such that $h(n) \geqslant M/\epsilon$ for each $n \geqslant n_1$. Then $\hat{g}(n) \leqslant \epsilon h(n)$ for each $n \geqslant n_1$. This implies that $\hat{g}(n) = o(h(n))$, as required. The case when $g(n) = O(h(n))$ can be shown similarly. ■

Since the Euler genus of a surface must be integer valued, we may also assume the genus function to be integer-valued. A similar assumption can be made in terms of parity in the orientable case. Often, we will be considering graphs on orientable surfaces. In this case the Euler genus of the surface is always even. We thus define for a genus function $g = g(n)$ the function $g^* = g^*(n)$ to be the function $g^*(n) = 2\lceil \hat{g}(n)/2 \rceil$. The function $g^*(n)$ is always even valued and satisfies $g(n) \leqslant g^*(n)$. Furthermore, it satisfies the following lemma.

Lemma 83. *Let $h = h(n)$ be non-decreasing and satisfy $h(n) \to \infty$ as $n \to \infty$. Given a real-valued function $g(n)$, the function $g^*(n)$ as defined above satisfies that if $g(n) = o(h(n))$ then $g^*(n) = o(h(n))$; and if $g(n) = O(h(n))$ then $g^*(n) = O(h(n)) \leqslant g(n) + 2$.*

Proof. A similar proof to Lemma 82 applies. ■

We can often assume without loss of generality, by considering $\hat{g}(n)$ or $g^*(n)$ instead of $g(n)$, that the genus function we consider is non-decreasing and in the orientable case also even-valued. In some cases we do however wish to keep these details, so in general work with $g(n)$ directly, but keep in mind that these assumptions can be made.

The outline of this section is as follows: We start by proving the following lower bounds on the size of the classes of graphs $\mathscr{G}_n^{g(n)}$ and $\mathscr{H}_n^{g(n)}$

$$\left|\mathscr{H}_n^{g(n)+1}\right| \geqslant \left|\mathscr{G}_n^{g(n)}\right| \geqslant n^{n-2} \cdot \left(\frac{n^2 - 3n}{2(n + g(n))}\right)^{g(n)} . \tag{3.15}$$

From these bounds, we can deduce that classes of graphs embeddable in an orientable or non-orientable surface of Euler genus any given $g(n) = \omega\left(\frac{n}{\log(n)}\right)$ are over-large. We go on to prove that the radius of convergence of the exponential generating function of the class of graphs embeddable in any orientable or non-orientable surface of Euler genus at most $g(n) = \Theta\left(\frac{n}{\log(n)}\right)$ is strictly smaller than the radius of convergence of the exponential generating function of the class of planar graphs $\rho_{\mathscr{P}} = \gamma_{\mathscr{P}}^{-1}$. In the process, we also find the following lower bounds on the sizes of $\mathscr{G}_n^{g(n)}$ and $\mathscr{H}_n^{g(n)}$ when $g(n) = o(n)$

$$\begin{aligned}\left|\mathscr{G}_n^{g(n)}\right| &\geqslant \gamma_{\mathscr{P}}^n \cdot n! \cdot n^{\left\lfloor \frac{g(n)}{2} \right\rfloor} \cdot (1 + o(1))^n \\ \left|\mathscr{H}_n^{g(n)}\right| &\geqslant \gamma_{\mathscr{P}}^n \cdot n! \cdot n^{\left\lfloor \frac{g(n)}{2} \right\rfloor} \cdot (1 + o(1))^n \ , \end{aligned} \tag{3.16}$$

which are better than the previous bounds when $g(n) = O\left(\frac{n}{\log(n)}\right)$.

In Section 3.3.2 we go on to find upper bounds on the sizes of both $\mathscr{G}_n^{g(n)}$ and $\mathscr{H}_n^{g(n)}$. By finding upper bounds for any $g(n)$ we deduce that in both the orientable and non-orientable case any given class of graphs of Euler genus at most $g(n) = O\left(\frac{n}{\log(n)}\right)$ is small. In Section 3.3.2.3 we find even tighter upper bounds for any genus $g(n) = o\left(\frac{n}{\log^3(n)}\right)$ from which we can deduce that the graph classes $\mathscr{G}_n^{g(n)}$ and $\mathscr{H}_n^{g(n)}$ have in fact a growth constant which is equal to the planar graph growth constant as long as $g(n) = o\left(\frac{n}{\log^3(n)}\right)$.

3.3.1 Lower Bounds

In this section, we find lower bounds on the size of the classes of graphs $\mathscr{G}_n^{g(n)}$ and $\mathscr{H}_n^{g(n)}$. For any genus function $g(n)$ such that $\limsup_{n\to\infty} \frac{g(n)\log(n)}{n} = \infty$ these are our best lower bounds, while for smaller genus we will improve these bounds in the later part of the section. From the original lower bounds, we will be able to deduce that for any genus $g(n)$ such that $\limsup_{n\to\infty} \frac{g(n)\log(n)}{n} = \infty$, the class of graphs embeddable in the orientable or non-orientable surface of Euler genus $g(n)$ is over-large. This also implies that we can only expect to have a growth constant for the graph classes $\mathscr{G}_n^{g(n)}$ and $\mathscr{H}_n^{g(n)}$ when $g(n)$ is $O\left(\frac{n}{\log(n)}\right)$.

In the second part of the section we find another lower bound on the size of $\mathscr{G}_n^{g(n)}$, which improves on the previous bound for any given $g(n) = O\left(\frac{n}{\log(n)}\right)$. We find bounds on the sizes of $\mathscr{G}_n^{g(n)}$ and $\mathscr{H}_n^{g(n)}$ below as in equation (3.16). From these bounds we can deduce that if $\mathscr{G}^{g(n)}$ has a growth constant for any given $g(n) = \Omega\left(\frac{n}{\log(n)}\right)$, the growth constant must be strictly bigger than the planar graph growth constant $\gamma_{\mathscr{P}}$. Note further that the same holds in the non-orientable case and can in fact even be deduced from the orientable case using Observation 66.

We now find lower bounds on the sizes of $\mathscr{G}_n^{g(n)}$ and $\mathscr{H}_n^{g(n)}$ which hold for any given $g(n)$. To do so, we first consider the orientable case, that is $\mathscr{G}_n^{g(n)}$, and deduce the non-orientable case from the orientable case using Observation 66.

Proposition 84. *Given a genus function $g(n)$ the following holds:*

$$\left|\mathscr{H}_n^{g(n)+1}\right| \geqslant \left|\mathscr{G}_n^{g(n)}\right| \geqslant n^{n-2} \cdot \left(\frac{n^2-3n}{2(n+g(n))}\right)^{g(n)} . \tag{3.17}$$

Further, given $c_0 > 0$ there exists a constant $c_1 > 0$ such that for $g(n) \leqslant c_0 \cdot n$ the following holds:

$$\left|\mathscr{H}_n^{g(n)+1}\right| \geqslant \left|\mathscr{G}_n^{g(n)}\right| \geqslant c_1^n \cdot n! \cdot n^{g(n)} . \tag{3.18}$$

Thus, as long as $\limsup_{n\to\infty} g(n) \cdot \frac{\log(n)}{n} = \infty$, the classes $\mathscr{H}_n^{g(n)}$ and $\mathscr{G}_n^{g(n)}$ are over-large.

Proof. We use the fact that Euler's formula (see Theorem 67) gives a lower bound on the number of edges a connected graph on n vertices needs to have in order to have Euler genus at least $g(n)+1$, as we have shown in Observation 72. By Observation 72 every cellular embedding of a connected graph in a surface whose Euler genus is at least $g(n)+1$ needs to satisfy

$$e \geqslant g(n) + n .$$

This implies that every connected graph with at most $g(n)+n-1$ edges can certainly be embedded in an orientable surface of Euler genus at most $g(n)$. In other words, every connected graph with at most $g(n)+n-1$ edges is contained in $\mathscr{G}_n^{g(n)}$. Furthermore, by Theorem 2.2 in [46], a fraction of at least $1/e$ graphs in $\mathscr{G}_n^{g(n)}$ is connected.

How many connected graphs on n vertices and with $g(n)+n-1$ edges are there? Pick a spanning tree on n vertices, there are n^{n-2} such trees, and then add any $g(n)$ of the $\binom{n}{2}-n+1$ potential edges that are not yet present in the graph. There are $\binom{\binom{n}{2}-n+1}{g(n)}$ choices for this. Thus, we construct $n^{n-2} \cdot \binom{\binom{n}{2}-n+1}{g(n)}$ connected graphs on $g(n)+n-1$ edges in total.

How often is each graph constructed? Each constructed graph has exactly $g(n)+n-1$ edges. We now guess the $g(n)$ extra edges that we need to delete in order to return to the original spanning tree. There are $\binom{g(n)+n-1}{g(n)}$ choices for this. Thus each graph is constructed (at most) $\binom{g(n)+n-1}{g(n)}$ times. Thus

$$\begin{aligned}\left|\mathscr{G}_n^{g(n)}\right| &\geqslant n^{n-2} \cdot \frac{\left(\frac{1}{2}(n^2-3n+2)\right)_{g(n)}}{(n+g(n)-1)_{g(n)}} \\ &\geqslant n^{n-2} \cdot \left(\frac{n^2-3n}{2(n+g(n))}\right)^{g(n)} ,\end{aligned} \tag{3.19}$$

as required for the first part of the proposition.

Given some constant $c_0 > 0$, we now suppose that $g(n) \leqslant c_0 \cdot n$. For n sufficiently large we then have

$$\begin{aligned}
\left|\mathscr{G}_n^{g(n)}\right| &\geqslant n^{n-2} \cdot \left(\frac{(n^2-3n+2)}{2(n+g(n))}\right)^{g(n)} \\
&\geqslant n^{n-2} \cdot \left(\frac{n^2}{4(n+g)}\right)^{g(n)} \\
&= n^{n-2} \cdot \left(\frac{n}{4}\right)^{g(n)} \cdot \left(\left(1+\frac{g}{n}\right)^{g(n)/n}\right)^{-n} \\
&\geqslant n^{n} \cdot \left(\frac{n}{4}\right)^{g(n)} \cdot e^{-n} \\
&\geqslant c_1^n \cdot n! \cdot n^{g(n)} \ ,
\end{aligned}$$

as required for the second part of the proposition.

In the orientable case we therefore have that as long as $g(n) = O(n)$ there is a constant $c_1 > 0$ such that

$$\left(\frac{\left|\mathscr{G}_n^{g(n)}\right|}{n!}\right)^{\frac{1}{n}} \geqslant c_1 \cdot e^{\frac{g(n)\log(n)}{n}} \ . \tag{3.20}$$

Note that the right hand side of equation (3.20) tends to infinity for any given $g(n)$ as long as $\omega(n/\log(n)) = g(n) = O(n)$.

As we mentioned previously, Observation 66 implies that the size of $\mathscr{G}_n^{g(n)-1}$ is a lower bound for the size of $\mathscr{H}_n^{g(n)}$. Now suppose that $\limsup_{n\to\infty} g(n) \cdot \frac{\log(n)}{n} = \infty$. By applying equation (3.20) to $\min\{g(n), n\}$, we see that $\mathscr{G}_n^{g(n)}$ and $\mathscr{H}_n^{g(n)}$ are over-large as required. ■

Hence, any class of graphs $\mathscr{G}_n^{g(n)}$ or $\mathscr{H}_n^{g(n)}$, for any given $g(n)$ such that $\limsup_{n\to\infty} g(n) \cdot \frac{\log(n)}{n} = \infty$, is over-large. Furthermore, the above bounds are our best lower bound on the size of $\mathscr{G}_n^{g(n)}$ or $\mathscr{H}_n^{g(n)}$ as long as $\limsup_{n\to\infty} g(n) \cdot \frac{\log(n)}{n} = \infty$. However, for Euler genus any given $g(n) = O\left(\frac{n}{\log(n)}\right)$ we find better bounds in the remainder of the section.

Proposition 85. *The radius of convergence ρ of the exponential generating function of $\mathscr{G}_n^{g(n)}$ or $\mathscr{H}_n^{g(n)}$ for any given $g(n) = \Omega\left(\frac{n}{\log(n)}\right)$ must be strictly smaller than $\rho_{\mathscr{P}} = \frac{1}{\gamma_{\mathscr{P}}}$, the radius of converge in the planar case. In other words, $\rho < \frac{1}{\gamma_{\mathscr{P}}}$. Furthermore, for any given $g(n)$ such that $g(n) = o(n)$ the following lower bounds hold*

$$\begin{aligned}
\left|\mathscr{G}_n^{g(n)}\right| &\geqslant \gamma_{\mathscr{P}}^n \cdot n! \cdot n^{\left\lfloor \frac{g(n)}{2} \right\rfloor} \cdot (1+o(1))^n \\
\left|\mathscr{H}_n^{g(n)}\right| &\geqslant \gamma_{\mathscr{P}}^n \cdot n! \cdot n^{\left\lfloor \frac{g(n)}{2} \right\rfloor} \cdot (1+o(1))^n \ .
\end{aligned} \tag{3.21}$$

Proof. By Observation 77 as long as $k = o(n)$ and n is sufficiently large it holds that $\left|\mathscr{A}_n^{k(n)+2}\right| \geqslant \frac{n}{7} \cdot \left|\mathscr{A}_n^{k(n)}\right|$, where $\mathscr{A}^{k(n)} = \mathscr{G}^{k(n)}$ or $\mathscr{A}^{k(n)} = \mathscr{H}^{k(n)}$. Applying this $\lfloor g(n)/2 \rfloor$ times to $\mathscr{A}_n^{g(n)}$ we obtain, since we assume that $g(n) = o(n)$, the following for n sufficiently large

$$\left|\mathscr{A}_n^{g(n)}\right| \geqslant \left(\frac{n}{7}\right)^{\lfloor g(n)/2 \rfloor} \cdot \left|\mathscr{P}_n\right| \ , \tag{3.22}$$

where we denote by $\mathscr{P}_n$ be the class of planar graphs on n vertices. We know from [34] that for some constant $0 < c_2 < 1$

$$|\mathscr{P}_n| \sim c_2 \cdot n^{-\frac{7}{2}} \cdot \gamma_{\mathscr{P}}^n \cdot n! \ . \tag{3.23}$$

Combining the two equations, we can deduce the following lower bound on the size of $\mathscr{A}_n^{g(n)}$ for sufficiently large n

$$\left|\mathscr{A}_n^{g(n)}\right| \geqslant n! \cdot \gamma_{\mathscr{P}}^n \cdot n^{-\frac{7}{2}} \cdot \left(\frac{n}{7}\right)^{\lfloor g(n)/2 \rfloor} \ , \tag{3.24}$$

which is equivalent to saying

$$\begin{aligned}
\left(\frac{\left|\mathscr{A}_n^{g(n)}\right|}{n!}\right)^{\frac{1}{n}} &\geqslant \left(\frac{1}{n!} \cdot n! \cdot \gamma_{\mathscr{P}}^n \cdot n^{-\frac{7}{2}} \cdot \left(\frac{n}{7}\right)^{\lfloor g(n)/2 \rfloor}\right)^{\frac{1}{n}} \\
&= \left(\frac{\gamma_{\mathscr{P}}^n \cdot n^{\lfloor g(n)/2 \rfloor} \cdot n^{-\frac{7}{2}}}{7^{\lfloor g(n)/2 \rfloor}}\right)^{\frac{1}{n}} \\
&= \gamma_{\mathscr{P}} \cdot \left(\frac{n}{7}\right)^{\lfloor \frac{g(n)}{2} \rfloor \cdot \frac{1}{n}} \cdot n^{-\frac{7}{2n}} \ .
\end{aligned}$$

This gives us the required result.

■

From these bounds it can be deduced that if there exists a growth constant in the case when $g(n) = \Omega\left(\frac{n}{\log(n)}\right)$, then this growth constant must be strictly bigger than the planar graph growth constant. This follows from observing that by equation (3.25) for any given Euler genus $g(n) = \frac{\mu \cdot n}{\log(n)}$ and some fixed $\mu > 0$ in the orientable case

$$\begin{aligned}
\left(\frac{\left|\mathscr{G}_n^{\frac{\mu \cdot n}{\log(n)}}\right|}{n!}\right)^{\frac{1}{n}} &\geqslant \gamma_{\mathscr{P}} \cdot \left(\frac{n}{7}\right)^{\left\lfloor \frac{\mu}{2\log(n)} \right\rfloor} \cdot n^{-\frac{7}{2n}} \\
&\to \gamma_{\mathscr{P}} \cdot e^{\frac{\mu}{2}} \\
&> \gamma_{\mathscr{P}} \ .
\end{aligned} \tag{3.25}$$

The equivalent result for the non-orientable case can be deduced by applying Observation 66. For any given Euler genus $g(n) = \frac{\mu \cdot n}{\log(n)}$ and some fixed $\mu > 0$ we have

$$\begin{aligned}\left(\frac{\left|\mathscr{H}_n^{\frac{\mu \cdot n}{\log(n)}}\right|}{n!}\right)^{\frac{1}{n}} &\geqslant \left(\frac{\left|\mathscr{G}_n^{\frac{\mu \cdot n}{\log(n)}-1}\right|}{n!}\right)^{\frac{1}{n}} \\ &\geqslant \gamma_{\mathscr{P}} \cdot \left(\frac{n}{7e}\right)^{\left\lfloor \frac{\mu}{2\log(n)} - \frac{1}{n} \right\rfloor} \cdot n^{-\frac{7}{2n}} \\ &\to \gamma_{\mathscr{P}} \cdot e^{\frac{\mu}{2}} \\ &> \gamma_{\mathscr{P}} \ . \end{aligned} \tag{3.26}$$

In conclusion, from Proposition 84 we know that as long as $\limsup_{n\to\infty} g(n) \cdot \frac{\log(n)}{n} = \infty$, the classes $\mathscr{H}_n^{g(n)}$ and $\mathscr{G}_n^{g(n)}$ are over-large, so their exponential generating functions have radius of convergence equal to zero. From Proposition 85 and equations (3.25) and (3.26) it can further be seen that if any of the graph classes $\mathscr{G}_n^{g(n)}$ or $\mathscr{H}_n^{g(n)}$ for any given $g(n) = \Omega\left(\frac{n}{\log(n)}\right)$ does have a growth constant, it must be strictly bigger than the planar graph growth constant $\gamma_{\mathscr{P}}$. In the next section we will show that the radius of convergence is also still strictly greater than zero in this case. Compared to the bounds in Proposition 84, Proposition 85 also provides us with improved lower bounds on the size of the class of graphs $\mathscr{G}_n^{g(n)}$ and $\mathscr{H}_n^{g(n)}$ for any given $g(n) = O\left(\frac{n}{\log(n)}\right)$.

3.3.2 Upper Bounds

After having established lower bounds on the sizes of $\mathscr{G}_n^{g(n)}$ and of $\mathscr{H}_n^{g(n)}$, we now go on to find upper bounds for the sizes of these classes. By doing so, we show that both $\mathscr{G}_n^{g(n)}$ and $\mathscr{H}_n^{g(n)}$ are small for any given $g(n) = O\left(\frac{n}{\log(n)}\right)$. The general idea of the proof of these bounds is to construct all possible graphs in these graph classes (plus additional non-simple graphs) by considering one-face embedded graphs and adding extra edges in the unique face of the embedding. We begin by defining what it means for an embedding to be a one-face embedding.

Definition. *We define a* ***one-face embedding*** *or* ***unicellular*** *embedding to be a cellular embedding of a connected graph that has just a single face.*

In the following section, Section 3.3.2.1, we will bound the size of $\mathscr{G}_n^{g(n)}$ for any given $g(n) = O\left(\frac{n}{\log(n)}\right)$ above by $c_3^{n+g(n)} \cdot n^{g(n)} \cdot n!$ for some constant $c_3 > 0$. This bound is the same up to some constant to the power n as the lower bound we found in the previous section, Section 3.3.1, as long as $g(n) = O(n)$. Note however that for $g(n) = \omega(n)$ the lower bound we found has leading term $\left(\frac{n^2}{g(n)}\right)^{g(n)}$, while the upper bound has leading term

$g(n)^{g(n)}$. We hope to bring these upper and lower bounds closer together in the future. By using a similar proof idea, we then find a similar upper bound to the non-orientable case by bounding the size of $\mathscr{H}_n^{g(n)}$ above by $c_7^{n+g(n)} \cdot n! \cdot (n+g(n))^{\frac{3g(n)}{2}} \cdot g(n)^{-\frac{g(n)}{2}} \cdot n!$ for some constant $c_7 > 0$. Note that it would be possible to deduce upper bounds on $\mathscr{G}_n^{g(n)}$ from bounds on $\mathscr{H}_n^{g(n)}$ using Observation 66. However, our results in the orientable case turn out to be more precise than in the non-orientable case. This is the case since an exact formula for the number of unicellular maps on orientable surfaces exist, while no such exact formula exists for unicellular maps on non-orientable surfaces. Hence, the non-orientable case is less precise and slightly more involved than the orientable case even though the basic idea of the proof is essentially the same. For this reason, we start with the orientable case.

3.3.2.1 Orientable Case

In the orientable case, we can find the following upper bounds on the size of the graph class containing all graphs on n vertices embeddable in the surface $\mathbf{S}_{g(n)/2}$.

Theorem 86. *For any fixed function $g(n)$, the class of graphs embeddable in an orientable surface of Euler genus at most $g(n)$ satisfies for n sufficiently large for some constant $c_3 > 0$ and any $g(n)$*

$$\left|\mathscr{G}_n^{g(n)}\right| \leqslant c_3^{n+g(n)} \cdot g(n)^{g(n)} \cdot n! \ .$$

This bound also immediately implies that the class of graphs $\mathscr{G}_n^{g(n)}$ is small for any $g(n) = O\left(\frac{n}{\log(n)}\right)$.

Proof. To establish this upper bound on the size of $\mathscr{G}_n^{g(n)}$ we consider cellular one face embeddings of connected graphs in $\mathbf{S}_{g(n)/2}$ and the addition of extra edges in the unique face of the embedding to construct more general graphs embeddable in the surface. This will allow us to find an upper bound on the number of graphs cellularly embeddable in the surface of Euler genus $g(n)$. Note however that $\mathscr{G}_n^{g(n)}$ also contains disconnected graphs and graphs that are cellularly embeddable only on surfaces of smaller genus than $g(n)$. However, by Lemma 78 we know that most connected graphs in $\mathscr{G}_n^{g(n)}$ are in fact cellularly embeddable in $\mathbf{S}_{g(n)/2}$, where by Lemma 83 we may assume that $g(n)$ is even. Further, by Theorem 2.2 in [46], a fraction of at least $1/e$ of all graphs in the graph class $\mathscr{G}_n^{g(n)}$ is connected. Since $1/e > 1/3$, Lemma 78 together with Theorem 2.2 in [46] implies that we count at least $1/3$ of all graphs in $\mathscr{G}_n^{g(n)}$ by counting all connected graphs with cellular embeddings in $\mathbf{S}_{g(n)/2}$ as long as n is sufficiently large.

We thus focus on counting connected graphs cellularly embeddable in $\mathbf{S}_{g(n)/2}$. Let G be a connected graph cellularly embeddable in $\mathbf{S}_{g(n)/2}$ and fix one of its cellular embeddings in this surface. Suppose further that this embedding has f faces say. We now make

use of Lemma 80 and Lemma 81. These lemmas state that it is possible to construct a unicellular subgraph W of G with $f-1$ fewer edges than G by repeatedly deleting edges of G that lie in the boundary of two faces. The induced embedding of W in $\mathbf{S}_{g(n)/2}$ is hence a connected cellular one-face embedding.

This implies that every unicellular embedding on $\mathbf{S}_{g(n)/2}$ with n vertices can be constructed by taking a one-face cellular embedded simple connected graph W on n vertices and adding edges in the unique face of its embedding.

We now find an upper bound on the number of graphs cellularly embeddable in $\mathbf{S}_{g(n)/2}$ by counting the number of such one-face embeddings of graphs W on n vertices and then finding an upper bound on the number of ways edges can be added in the face. Note that by the fact that the embedding is cellular, the open face is homeomorphic to a disc.

The number of unicellular maps rooted at an oriented edge in $\mathbf{S}_{g(n)/2}$ on n unlabelled vertices is equal by [55] to

$$\frac{(2n+2g(n)-2)!}{2^{g(n)}\cdot n!\cdot(n+g(n)-1)!}\cdot\sum_{\substack{i_1+\cdots+i_n=g(n)\\ i_1,\ldots,i_n\geqslant 0}}\prod_{j=1}^{n}\frac{1}{2i_j+1}\ . \tag{3.27}$$

We now try to bound on equation (3.27) above. To do so we make use of the fact that

$$\sum_{\substack{i_1+\cdots+i_n=g(n)\\ i_1,\ldots,i_n\geqslant 0}} 1=\binom{n+g}{n}$$

and

$$\prod_{j=1}^{n}\frac{1}{2i_j+1}\leqslant\frac{1}{2g(n)+1}\leqslant 1\ .$$

Equation (3.27) can thus be bounded above for all $g(n)$ by

$$\begin{aligned}
&\frac{(2n+2g(n)-2)!}{2^{g(n)}\cdot n!\cdot(n+g(n)-1)!}\cdot\binom{n+g(n)}{n}\\
&=\binom{2n+2g(n)-2}{n+g(n)-1}\cdot\frac{(n+g(n)-1)!}{n!}\cdot 2^{-g(n)}\cdot\binom{n+g(n)}{n}\\
&\leqslant 2^{2n+2g(n)-2}\cdot 2^{-g(n)}\cdot\binom{n+g(n)}{n}^2\cdot g(n)!\\
&\leqslant 2^{2n+g(n)}\cdot 2^{2n+2g(n)}\cdot g(n)!\\
&=2^{4n+3g(n)}\cdot g(n)!\\
&\leqslant 2^{4n+3g(n)}\cdot g(n)^{g(n)}\ .
\end{aligned}\tag{3.28}$$

The number of possible cellular one-face, with the facial walk starting along a given oriented edge, labeled graph embeddings in $\mathbf{S}_{g(n)/2}$ on n vertices is bounded above by $n!$ times the number of unicellular maps on $\mathbf{S}_{g(n)/2}$ on n vertices. Equation (3.28) thus

gives, when multiplied by $n!$, an upper bound on the number of one-face embeddings W of graphs on vertex set $[n]$ on the surface $\mathbf{S}_{g(n)/2}$.

Once we have found a unicellularly embedded graph W, we need to add edges in the interior of the face bounded by the facial walk of the unique face of W such that no two edges in the interior cross. Recall that the embedding of W is cellular, so the unique face is homeomorphic to a disc. There are $2g(n) + 2n - 2$ vertices in the boundary walk around the unique face, including repetitions, since every edge of W appears exactly twice in the boundary walk. From [32] we know that for any k there are at most $(2 + 3\sqrt{2})^k$ dissections of an k-gon, where the vertices are labelled $v_1, \ldots, v_k$ clockwise around the cycle on k vertices for k sufficiently large. For a given W, allowing double edges and loops, there are hence at most $(2 + 3\sqrt{2})^{2g(n)+2n-2}$ possible ways of filling the face, where $k = 2g(n) + 2n - 2$. This can be seen by taking the rooted edge of W as the starting edge of the boundary cycle and noting that the labels of the remaining vertices in the boundary cycle are determined by the embedding of W. Note that this bound is certainly not tight since vertices appear more than once in the boundary cycle. This implies that some of these dissections of the face will contain double edges or even self loops. Putting all this together, we have in total, for some constant $c_3 > 0$, at most

$$(2 + 3\sqrt{2})^{2g(n)+2n} \cdot 2^{4n+3g(n)} \cdot g(n)^{g(n)} \cdot n! \ll c_3^{n+g(n)} \cdot g(n)^{g(n)} \cdot n! \tag{3.29}$$

cellular embeddings of connected simple graphs on $\mathbf{S}_{g(n)/2}$ on n vertices and hence also at most this many connected simple graphs which have an cellular embedding in $\mathbf{S}_{g(n)/2}$.

By Theorem 56 a fraction of at least $1/e$ of all graphs in $\mathscr{G}_n^{g(n)}$ is connected. Furthermore, each connected graph in $\mathscr{G}_n^{g(n)}$ has, as long as $g(n) = o(n^2)$, with high probability a cellular embedding in $\mathbf{S}_{g(n)/2}$. Since $1/e < 1/3$, this implies that for some constant $c_3 > 0$ and sufficiently large n we have

$$\begin{aligned}\left|\mathscr{G}_n^{g(n)}\right| &\leqslant 3 \cdot \text{ number of cellular embeddings of connected } n\text{-vertex graphs on } \mathbf{S}_{\frac{g(n)}{2}} \\ &\ll c_3^{n+g(n)} \cdot g(n)^{g(n)} \cdot n! \ ,\end{aligned} \tag{3.30}$$

as required. ■

Equation (3.29) gives us an upper bound on the size of $\mathscr{G}_n^{g(n)}$. Note that in Section 3.3.1 equations (3.20) and (3.25) we have found lower bounds on the size of $\mathscr{G}_n^{g(n)}$. Combining these with the upper bound we found in this section, we have the following results:

- *Range* $g(n) = O\left(\frac{n}{\log(n)}\right)$
  In this range there exists a constant $c_3 > 0$ such that for sufficiently large n the following holds

$$n! \cdot \gamma_{\mathscr{P}}^n \cdot n^{\lfloor g(n)/2 \rfloor} \leqslant \left|\mathscr{G}_n^{g(n)}\right| \leqslant n! \cdot c_3^{n+g(n)} \cdot g(n)^{g(n)} \ . \tag{3.31}$$

- *Range $g(n) \leqslant \alpha n$ for some $\alpha > 0$*
  In this range there exist constants $c_1 > 0$ and $c_3 > 0$ depending on α such that for sufficiently large n the following holds

$$n! \cdot c_1^n \cdot n^{g(n)} \leqslant \left|\mathscr{G}_n^{g(n)}\right| \leqslant n! \cdot c_3^{n+g(n)} \cdot g(n)^{g(n)} \ . \tag{3.32}$$

- *Range $g(n) = \omega(n)$*
  In this range there exist constants $c_4 > 0$ and $c_3 > 0$ such that for sufficiently large n the following holds

$$c_4^{g(n)} \cdot n^n \cdot \left(\frac{n^2}{g(n)}\right)^{g(n)} \leqslant \left|\mathscr{G}_n^{g(n)}\right| \leqslant n! \cdot c_3^{n+g(n)} \cdot g(n)^{g(n)} \ . \tag{3.33}$$

3.3.2.2 Non-orientable Case

As mentioned previously, the main idea of the proof for the non-orientable case is essentially the same as the proof in the orientable case. However, there is no known upper bound on the number of unicellular maps on non-orientable surfaces for non-constant genus as in the orientable case, which makes the proof slightly more involved as we first need to find such a bound. We find this bound by relating cubic unicellular maps, for which an exact counting formula is known, to general unicellular maps.

Theorem 87. *For any given function $g(n) = O\left(\frac{n}{\log(n)}\right)$, the class of graphs $\mathscr{H}_n^{g(n)}$ is small. Furthermore there exists a constant $c_7 > 0$ such that for sufficiently large n*

$$\left|\mathscr{H}_n^{g(n)}\right| \leqslant c_7^{n+g(n)} \cdot n! \cdot g(n)^{-\frac{g(n)}{2}} \cdot (g(n)+n)^{\frac{3g(n)}{2}} \ . \tag{3.34}$$

Proof. To establish this upper bound on the size of $\mathscr{H}_n^{g(n)}$, we consider cellular embeddings of connected graphs in $\mathbf{N}_{g(n)}$ and the addition of extra edges in the unique face to construct more general connected graphs embeddable in the surface. This will allow us, as in the orientable case, to find an upper bound on the number of graphs cellularly embeddable in $\mathbf{N}_{g(n)}$. Note however that like in the orientable case $\mathscr{H}_n^{g(n)}$ also contains graphs cellularly embeddable in surfaces of smaller Euler genus. As in the orientable case, we can reduce the number of surfaces we need to consider using Lemma 79.

Lemma 79 implies that with high probability a connected graph chosen uniformly at random from $\mathscr{H}_n^{g(n)}$ has with high probability a cellular embedding in $\mathbf{N}_{g(n)}$ or $\mathbf{N}_{g(n)-1}$, as long as $g(n) = o(n^2)$. Note furthermore that by Theorem 2.2 in [46], a fraction of at least $1/e$ of all graphs in the graph class $\mathscr{H}_n^{g(n)}$ are connected. This implies that we count at least $1/3$, since $1/3 > 1/e$, of all graphs in $\mathscr{H}_n^{g(n)}$ by counting all connected graphs with cellular embeddings in $\mathbf{N}_{g(n)}$ or $\mathbf{N}_{g(n)-1}$ for n sufficiently large. Thus, let G be a connected graph cellularly embeddable in $\mathbf{N}_{g(n)}$ or $\mathbf{N}_{g(n)-1}$ and fix one of its cellular

embeddings in this surfaces. Suppose further that this embedding has f faces. We now make use of Lemma 80 and Lemma 81. These lemmas state that it is possible to construct a unicellular subgraph W of G with $f-1$ fewer edges than G by repeatedly deleting edges of G that lie in the boundary of two faces. The induced embedding of W in $\mathbf{N}_{g(n)}$ or $\mathbf{N}_{g(n)-1}$ is hence a connected cellular one-face embedding.

As a result, every simple, connected, cellular graph embedding on $\mathbf{N}_{g(n)}$ on n vertices can be constructed by taking a one-face connected embedded simple graph W on n vertices and adding edges in the unique face of its embedding.

We now find an upper bound on the number of connected graphs cellularly embeddable in $\mathbf{N}_{g(n)}$ or $\mathbf{N}_{g(n)-1}$ or $\mathbf{S}_{\lfloor g(n)/2 \rfloor}$ on n vertices by counting the number of such one-face embeddings of graphs W on n vertices and then finding an upper bound on the number of ways edges can be added in the unique face. Note that the open face is homeomorphic to a disc since we are looking at cellular embeddings.

In the orientable case, an exact counting formula for the number of n-vertex unicellular maps on surfaces exists which gives us, multiplied by $n!$, an upper bound on the number of connected unicellular embedded graphs. To our knowledge this is not the case for non-orientable surfaces at the time of writing and only an asymptotic bound for fixed non-orientable genus surfaces exists [23], which is not applicable in our case. This makes obtaining an upper bound on the number of connected unicellular embedded graphs in a non-orientable surface more involved. We bound the number of labeled connected unicellular maps with n vertices on $\mathbf{N}_{g(n)}$ by relating general connected labeled unicellular maps to unlabeled pre-cubic (so only containing vertices of degree one and three) connected unicellular maps on non-orientable surfaces, since an exact counting formula exists for these. Taking any labeled connected unicellular map on n vertices, we can construct a pre-cubic unicellular map, possibly with parallel edges and self-loops, from it as follows: By replacing all vertices of degree two by an edge we obtain a map with no vertices of degree two. We now remove the labels from the vertices. By repeatedly splitting vertices of degree greater than three into two adjacent vertices with the original neighbours adjacent to either one of the new vertices in the natural way and such that the degree of each new vertex is at least three (see Figure 3.3.2.2), we obtain a unicellular pre-cubic map (or a single loop) assuming that the original graph has more than one vertex. Since obtaining a single loop is however only possible in the case when $g(n) = 0$, this issue is negligible for larger $g(n)$ and can be avoided by just adding one to the number of pre-cubic unicellular maps we construct our general unicellular maps from.

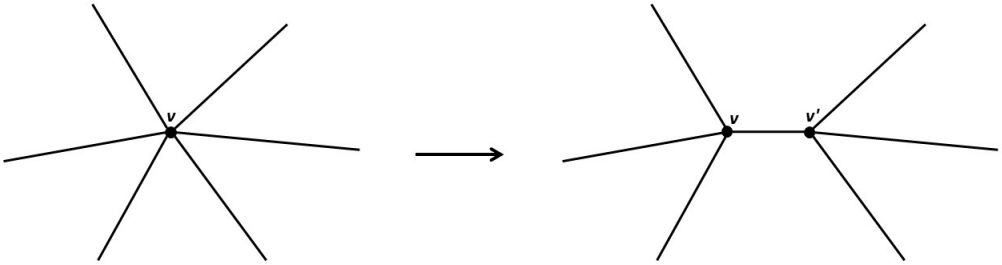


Figure 3.10: Splitting a vertex v of degree greater than three

This implies on the other hand that we can construct any connected labeled unicellular map on a surface by contracting edges in a connected pre-cubic unicellular map, labeling the vertices and then adding vertices in the middle of an edge.

An exact counting formula for the number of unlabeled connected pre-cubic (so all vertices of degree 1 or 3) unicellular maps on a surface is known by [23]. We now use this to find an upper bound on the number of general labeled unicellular maps by connecting them in the above mentioned way. Suppose the number of vertices of degree two we later add in the graph is equal to l, where $l < n$. We do not know the exact number of edges in our pre-cubic maps, however the maximum number of edges we could have is $3(g(n) + n - l)$. Note that this is probably not a tight bound, however sufficiently good for our purpose. Further, by deleting vertices of degree two from a simple graph, we can construct double edges and even self loops, so our pre-cubic unicellular maps are not necessarily simple.

We obtain the bound on the maximum number of edges as follows: Note that once we have replaced the l vertices of degree two, we are left with a unicellular graph on $n' = n-l$ vertices. By Euler's formula (see Theorem 67) and since the graph is unicellular, this graph has exactly $g(n) + n' - 1$ edges. In the next step of constructing the pre-cubic graph we split each vertex of degree greater than three until no such vertices are left. In every splitting step we decrease the degree of a vertex and add a new vertex as well as a new edge. To obtain vertices all of degree three from a vertex of degree $d(v) > 3$ we need to make exactly $d(v) - 3$ vertex splits. In total, summing over all vertices, we thus need to make exactly

$$\sum_{\substack{v \in V(G) \\ d(v) > 3}} (d(v) - 3)$$

splits to turn the unicellular graph on n' vertices into a pre-cubic unicellular graph, where $V(G)$ denotes the set of vertices of G, the unicellular graph with no degree two vertices, and $d(v)$ denotes the degree of the vertex v. The number of splits can be bounded above by

$$\begin{aligned}\sum_{\substack{v\in V(G)\\ d(v)>3}} (d(v)-3) &\leqslant \sum_{\substack{v\in V(G)\\ d(v)>3}} d(v)\\ &\leqslant \sum_{v\in V(G)} d(v)\\ &\leqslant 2e(G)\ , \end{aligned} \tag{3.35}$$

where $e(G)$ denotes the number of edges of G. The pre-cubic graph we construct has hence at most $3e(G) = 3(g(n)+n'-1) < 3(g(n)+n-l)$ edges.

This gives us the following bounds on the number of pre-cubic graphs, that is before adding vertices of degree two. Note that for pre-cubic graphs the genus determines the parity of the number of edges present in the graph. Thus, if the number of edges is odd, Corollary 8 of [23] gives a formula for the number of non-orientable pre-cubic unicellular maps on a surface of genus $g(n)$ with $2m+1$ edges. Since this formula is monotonic increasing with the number of edges, we can find an upper bound by taking $2m+1 = 3(g(n)+n-l)$. For clarity, we start off by using h and m as in [23] before substituting in their values for them. We further use the notation $n' = n-l$. The number of non-orientable pre-cubic unicellular maps on a surface with $3(g(n)+n-l)$ edges, when $3(g(n)+n-l)$ is odd, is equal to

$$\begin{aligned}&3\cdot 2^{3h-2}\,\frac{h!}{(2h)!}\sum_{l=0}^{h-1}\left(\binom{2l}{l}16^{-l}\right)\,\frac{(2m)!}{6^h\,m!(m+1-3h)!}\\
&\leqslant 3\cdot 2^{3h-2}\binom{2h}{h}^{-1}\frac{1}{h!}\,2^{-2h}\binom{2m}{m}\,\frac{m!}{6^h\,(m+1-3h)!}\\
&\leqslant 3^{1-h}\,2^{-2}\,2^{-2h}\,(h!)^{-1}\,2^{2m}\,m!\,((m+1-3h)!)^{-1}\\
&\leqslant 3^{1-h}\,2^{2m-2h-2}\,(h^h\,e^{-h})^{-1}\,(m^m\,e^{-m})\,((m+1-3h)^{m+1-3h}\,3^{-m-1+3h})^{-1}\\
&\leqslant 3^{1-h}\,2^{2m-2h-2}\,e^{1-2h}\,h^{-h}\,m^m\,(m-3h)^{-m+3h}\\
&\leqslant 3^{1-\frac{g(n)}{2}}\,2^{2g(n)+3n'-3}\,e^{1-g(n)}\,\left(\frac{g(n)}{2}\right)^{-\frac{g(n)}{2}}\,\left(\frac{3(g(n)+n')}{2}\right)^{\frac{3(g(n)+n')}{2}}\,\left(\frac{3n'}{2}\right)^{-\frac{3n'}{2}}\\
&\leqslant 3^{1+g(n)}\,2^{3g(n)+3n'-3}\,e^{1-g(n)}\,g(n)^{-\frac{g(n)}{2}}\,(g(n)+n')^{\frac{3(g(n)+n')}{2}}\,(n')^{-\frac{3n'}{2}}\\
&\ll c_5^{n+g(n)}\,g(n)^{-\frac{g(n)}{2}}\,(g(n)+n-l)^{\frac{3g(n)}{2}}\ ,\end{aligned}$$

for some constant $c_5 > 0$ which we choose sufficiently large to also satisfy the case when the number of edges is even which follows in the next equation. Similarly if the

actual number of edges is even Corollary 9 of [23] gives a formula for the number of non-orientable pre-cubic unicellular maps on a surface of genus $g(n)$ with $2m$ edges. Since this formula is monotonic in the number of edges, we can find an upper bound by taking $2m = 3(g(n) + n - l)$. As in the odd case, we start off by using h and m as in [23] for clarity before substituting in their values. The number of non-orientable pre-cubic unicellular maps on a surface with $3(g(n) + n - l)$ edges, when $3(g(n) + n - l)$ is even, is equal to

$$\begin{aligned}
&\frac{4^{m+\lfloor h\rfloor-1}\,(m-1)!}{6^{\lfloor h\rfloor}\,(2h-1)!!\,(m-1-3\lfloor h\rfloor)!}\\
&\leqslant 4^{m+\lfloor h\rfloor-1}\,6^{-\lfloor h\rfloor}\,\frac{m!}{2^{h-1/2}\,(h-1/2)!\,(m-1-3\lfloor h\rfloor)!}\\
&\leqslant 2^{2m-3/2}\,3^{-\lfloor h\rfloor}\,\frac{m^m\,e^{-m}}{h^{h-1/2}\,e^{-h+1/2}\,(m-1-3\lfloor h\rfloor)^{m-1-3\lfloor h\rfloor}\,e^{-m+3\lfloor h\rfloor+1}}\\
&\leqslant 2^{2m-3/2}\,3^{-\lfloor h\rfloor}\,e^{-m-2h}\,m^m\,h^{-h+1/2}\,(m-1-3\lfloor h\rfloor)^{-m+1+3\lfloor h\rfloor}\\
&\leqslant 2^{3(g(n)+n')}\,3^{-\frac{g(n)-1}{2}}\,e^{-\frac{5g(n)-3n'}{2}}\,\left(\frac{3(g(n)+n')}{2}\right)^{\frac{3(g(n)+n')}{2}}\,\left(\frac{g(n)}{2}\right)^{\frac{-g(n)+1}{2}}\,\left(\frac{3n'}{2}\right)^{-\frac{3n'}{2}}\\
&\leqslant 2^{2g(n)+3n'+\frac{1}{2}}\,3^{g(n)}\,e^{-\frac{5g(n)-3n'}{2}}\,(g(n)+n')^{\frac{3(g(n)+n')}{2}}\,g(n)^{-\frac{g(n)-1}{2}}\,(n')^{-\frac{3n'}{2}}\\
&\ll c_5^{n+g(n)}\,g(n)^{-\frac{g(n)}{2}}\,(g(n)+n-l)^{\frac{3g(n)}{2}}\ .
\end{aligned}$$

So in each case the bound is of order $c_5^{n+g(n)}\,g(n)^{-g(n)/2}\,(3g(n)+n-l)^{3g(n)/2-1}$. Note further, that we need to account for the different number of edges the pre-cubic unicellular map can have. We know that it has at most $3(g(n) + n - l)$ edges and the counting formulas are increasing with the number of edges. As a result, there are at most $3(g(n) + n - l)$ times as many pre-cubic unicellular maps with at most $3(g(n) + n - l)$ edges than there are pre-cubic unicellular maps with exactly $3(g(n) + n - l)$ edges. The number of connected pre-cubic unicellular maps with at most $3(g(n) + n - l)$ edges is thus at most $c_5^{n+g(n)}\,g(n)^{-g(n)/2}\,(3g(n)+n-l)^{3g(n)/2}$.

We can obtain any general connected unicellular map with no vertices of degree two from some connected pre-cubic unicellular map by contracting some number of edges. Note that the order in which we contract these edges has no influence on the final graph we obtain. We already noted that the graphs we counted above have at most $3(g(n)+n-l)$ edges. There are thus $2^{3(g(n)+n-l)}$ possible subsets of edges to contract. We can hence construct from a given connected unicellular pre-cubic map at most $2^{3(g(n)+n-l)}$ unicellular maps with no degree two vertices on the same surface.

We now label this map. Note that we will eventually want to have a map on the vertex set $[n]$, so there are n choices for the label of the first vertex, $n-1$ for the second and so

on. Since we have not added the l vertices of degree two so far, our graph currently has $n-l$ vertices. To label these, we thus have $(n)_{n-l}$ choices. We assumed that our general unicellular graph contains $l < n$ vertices of degree two. We now add these in the middle of edges of our labeled unicellular map containing no vertices of degree two. Note that the order in which we add the vertices of degree two to the graph does not matter, so we choose to add them in order of increasing label. By Euler's formula (see Theorem 67) there are at most $n+g(n)-1$ edges in a connected one-face embedded graph, so at most $n+g(n)-1$ choices of where we can add a vertex of degree two in the middle of an edge. So we have at most $(n+g-1)^l$ choices of where to add the l vertices of degree two. Thus, there are at most

$$\sum_{l=0}^{n-1} c_5^{n+g(n)}\, g(n)^{-\frac{g(n)}{2}}\, (g(n)+n-l)^{\frac{3g(n)}{2}}\, 2^{3(g(n)+n-l)}\, (n)_{n-l}\, (g(n)+n-1)^l \tag{3.36}$$

labeled one-face graph embeddings W on $\mathbf{N}_{g(n)}$ on n vertices. We can bound this above by

$$\begin{aligned}
&\sum_{l=0}^{n-1} c_5^{n+g(n)}\, g(n)^{-\frac{g(n)}{2}}\, (g(n)+n-l)^{\frac{3g(n)}{2}}\, 2^{3(g(n)+n-l)}\, \frac{en\left(\frac{n}{e}\right)^n}{e\left(\frac{l}{e}\right)^l}\, (g(n)+n-1)^l \\
&\leqslant \sum_{l=0}^{n-1} c_5^{n+g(n)}\, 2^{3(g(n)+n-l)}\, e^{l-n}\, g(n)^{-\frac{g(n)}{2}}\, (g(n)+n-l)^{\frac{3g(n)}{2}}\, n^{n+1}\, l^{-l}\, (g(n)+n-1)^l \\
&\leqslant c_6^{n+g(n)}\, n!\; g(n)^{-\frac{g(n)}{2}}\, (g(n)+n)^{\frac{3g(n)}{2}}\ .
\end{aligned}$$

Once we have found W, a unicellular graph embedding, we need to add edges in the interior of the face bounded by the facial walk of the unique face of W such that no two edges in the interior cross. There are $2g+2n-2$ vertices in the boundary walk around the unique face, including repetitions, since every edge of W appears exactly twice in the boundary walk. From [32] we know that for any k there are at most $(2+3\sqrt{2})^k$ dissections of an k-gon, where the vertices are labelled $v_1,\ldots,v_k$ clockwise around the cycle on k vertices for k sufficiently large. Note that unlike in the orientable case, the unicellular map is not rooted at an edge, meaning we have to pick one of the $2g(n)+n-2$ vertices as the starting vertex and fix this vertex. This does hence not affect our number of choices. We do however need to assign a sign to each edge, giving us an extra factor of $2^{2g(n)+2n}$. For a given W, allowing double edges and loops, there are hence at most $2^{2g(n)+2n}\cdot(2+3\sqrt{2})^{2g(n)+2n-2}$ possible ways of filling the face, where $k=2g(n)+2n-2$. Note that this bound is certainly not tight since vertices appear more than once in the boundary cycle. This implies that some of these dissections of the face will contain double edges or even self loops. In total there are hence at most

$$\begin{aligned}
&2^{2g(n)+2n}\,(3+2\sqrt{2})^{2g(n)+2n-2}\, c_6^{n+g(n)}\, n!\; g(n)^{-\frac{g(n)}{2}}\, (g(n)+n)^{\frac{3g(n)}{2}} \\
&\qquad\ll c_7^{n+g(n)}\, n!\; g(n)^{-\frac{g(n)}{2}}\, (g(n)+n)^{\frac{3g(n)}{2}}
\end{aligned} \tag{3.37}$$

cellular embeddings of simple connected graphs in $\mathbf{N}_{g(n)}$ on n vertices, that is at most this many connected simple graphs that have a cellular embedding in $\mathbf{N}_{g(n)}$.

By Theorem 2.2 in [46], at least $1/e$ of all graphs in the graph class $\mathscr{H}_n^{g(n)}$ are connected. Further, by Lemma 79 each connected graph in $\mathscr{H}_n^{g(n)}$ has, as long as $g(n) = o(n^2)$, with high probability an embedding in either $\mathbf{N}_{g(n)}$ or $\mathbf{N}_{g(n)-1}$ or $\mathbf{S}_{\lfloor g(n)/2 \rfloor}$. Note that we previously counted the number of graphs cellularly embeddable in $\mathbf{S}_{\lfloor g(n)/2 \rfloor}$ in the previous section. We just add this on to the bound that we have obtained for graphs embeddable in $\mathbf{N}_{g(n)}$ and $\mathbf{N}_{g(n)-1}$ to obtain the desired result. Note that the bounds on $\mathbf{S}_{\lfloor g(n)/2 \rfloor}$ are of the same leading terms as those for $\mathbf{N}_{g(n)}$, so adding on the orientable case does not change the bound as long as c_7 is chosen sufficiently large. Since $1/e < 1/3$ we have counted in the above, for n large enough, at least $1/6$ of all graphs embeddable in $\mathbf{N}_{g(n)}$ or $\mathbf{N}_{g(n)-1}$. Thus for some constant $c_7 > 0$ and when n is sufficiently large

$$\left|\mathscr{H}_n^{g(n)}\right| \leqslant c_7^{n+g(n)} \cdot n! \cdot g(n)^{-\frac{g(n)}{2}} \cdot (g(n)+n)^{\frac{3g(n)}{2}} \ , \tag{3.38}$$

as required. ■

As a result, the class of graphs embeddable in the non-orientable surface of Euler genus at most any given $g(n) = O\left(\frac{n}{\log(n)}\right)$ is small, since $\mathscr{H}_n^{g(n)}$ contains for n sufficiently large at most

$$c_7^{n+g(n)} \cdot n! \cdot g(n)^{-\frac{g(n)}{2}} \cdot (g(n)+n)^{\frac{3g(n)}{2}} = \Theta\left(n^{g(n)} \cdot n!\right) \tag{3.39}$$

elements.

Equation (3.38) gives us an upper bounds on the size of $\mathscr{H}_n^{g(n)}$. In the previous section, Section 3.3.1, we further found lower bounds on the size of $\mathscr{H}_n^{g(n)}$. Putting these together, we see that

- *Range* $g(n) = O\left(\frac{n}{\log(n)}\right)$
  In this range there exists a constant $c_7 > 0$ such that for sufficiently large n the following holds

$$n! \cdot \gamma_{\mathscr{P}}^n \cdot n^{\frac{g(n)}{2}} \leqslant \left|\mathscr{H}_n^{g(n)}\right| \leqslant c_7^{n+g(n)} \cdot n! \cdot g(n)^{-\frac{g(n)}{2}} \cdot (g(n)+n)^{\frac{3g(n)}{2}} \ . \tag{3.40}$$

- *Range* $\omega\left(\frac{n}{\log(n)}\right) = g(n) = o(n)$
  In this range there exist constants $c_1 > 0$ and $c_7 > 0$ such that for sufficiently large n the following holds

$$n! \cdot c_1^n \cdot n^{g(n)} \leqslant \left|\mathscr{H}_n^{g(n)}\right| \leqslant c_7^{n+g(n)} \cdot n! \cdot g(n)^{-\frac{g(n)}{2}} \cdot (g(n)+n)^{\frac{3g(n)}{2}} \ . \tag{3.41}$$

- *Range $g(n) = \alpha n$ for some $\alpha > 0$*
  In this range there exist constants $c_1 > 0$ and $c_7 > 0$ depending on α such that for sufficiently large n the following holds

$$n! \cdot c_1^n \cdot n^{\alpha n} \leqslant \left|\mathscr{H}_n^{g(n)}\right| \leqslant c_7^{n+g(n)} \cdot n! \cdot g(n)^{-\frac{g(n)}{2}} \cdot (g(n)+n)^{\frac{3g(n)}{2}} \ . \tag{3.42}$$

- *Range $g(n) = \omega(n)$*
  In this range there exist constants $c_4 > 0$ and $c_7 > 0$ such that for sufficiently large n the following holds

$$c_4^{g(n)} \cdot n^n \cdot \left(\frac{n^2}{g(n)}\right)^{g(n)} \leqslant \left|\mathscr{H}_n^{g(n)}\right| \leqslant c_7^{n+g(n)} \cdot n! \cdot g(n)^{-\frac{g(n)}{2}} \cdot (g(n)+n)^{\frac{3g(n)}{2}} \ . \tag{3.43}$$

3.3.2.3 Growth Constant up to Genus $g(n) = o\left(\frac{n}{\log^3(n)}\right)$

We aim to show that for any $g(n) = o\left(n/\log^3(n)\right)$, the class of graphs $\{\mathscr{G}_n^{g(n)}\}_{n\in\mathbb{N}}$ has a growth constant, which is in fact equal to the planar graph growth constant.

To show that $\mathscr{G}_n^{g(n)}$ has a growth constant for all $g(n) = o\left(n/\log^3(n)\right)$, we first show that a large portion of graph classes in $\{\mathscr{G}_n^{g(n)}\}_{n\in\mathbb{N}}$ has with high probability linearly many pendant vertices and use this result to bound the size of said graph class. To do so, we start by showing that in most cases, when going from n to $n+1$ vertices, the size of $\mathscr{G}_{n+1}^{g(n+1)}$ is not much bigger than the size of $\mathscr{G}_n^{g(n)}$. We formalise this in the following lemma.

Lemma 88. *Let $\epsilon > 0$ and let $g(n) = o\left(\frac{n}{\log(n)}\right)$ be any fixed non-decreasing, non-negative function. There exists a constant $c_\epsilon > 0$ such that for all n at most ϵn of all $m \in \{1, \cdots, n\}$ satisfy*

$$\left|\mathscr{G}_{m+1}^{g(m+1)}\right| \geqslant c_\epsilon \cdot (m+1) \cdot \left|\mathscr{G}_m^{g(m)}\right| \ . \tag{3.44}$$

Proof. Let $c_\epsilon = (3c_3)^{1/(1-\epsilon)}$. We now show that c_ϵ satisfies

$$\left|\mathscr{G}_{m+1}^{g(m+1)}\right| \geqslant c_\epsilon \cdot (m+1) \cdot \left|\mathscr{G}_m^{g(m)}\right| \ , \tag{3.45}$$

which would then complete the proof. Suppose for a contradiction that there are at least ϵn integers $m \in \{1, \cdots, n\}$ such that

$$\left|\mathscr{G}_{m+1}^{g(m+1)}\right| \geqslant (m+1) \cdot c_\epsilon \cdot \left|\mathscr{G}_m^{g(m)}\right| \ . \tag{3.46}$$

By iterating this and noting that by Lemma 76 for all $m \in \mathbb{N}$

$$\left|\mathscr{G}_{m+1}^{g(m+1)}\right| \geqslant (m+1) \cdot \left|\mathscr{G}_m^{g(m)}\right| \tag{3.47}$$

holds, we obtain that there exits an n such that

$$\begin{aligned}\left|\mathscr{G}_n^{g(n)}\right| &\geqslant n! \cdot c_\epsilon^{(1-\epsilon)n} \cdot g(n)^{g(n)/2} \\ &= n! \cdot 3^n \cdot c_3^n \cdot g(n)^{g(n)/2} \ . \end{aligned} \tag{3.48}$$

This contradicts, as long as $g(n) = o\left(\frac{n}{\log(n)}\right)$, the previously established result from equation (3.30) that

$$\left|\mathscr{G}_n^{g(n)}\right| \leqslant n! \cdot c_3^{n+g(n)} \cdot g(n)^{g(n)} \leqslant n! \cdot (2c_3)^n \ . \tag{3.49}$$

It follows that for all $\epsilon > 0$, $c_\epsilon = (3c_3)^{1/(1-\epsilon)}$ is such that at most ϵn of all $m \in \{1, \cdots, n\}$ satisfy

$$\left|\mathscr{G}_{m+1}^{g(m+1)}\right| \geqslant c_\epsilon \cdot (m+1) \cdot \left|\mathscr{G}_m^{g(m)}\right| \ , \tag{3.50}$$

as required. ■

Fix $\epsilon > 0$. Now denote by $\mathcal{S}_\epsilon$ the set of all $m \in \mathbb{N}$ such that going from m to $m+1$ is not a "jump step", so $g(m) = g(m+1)$, and such that

$$\left|\mathscr{G}_{m+1}^{g(m+1)}\right| \leqslant c_\epsilon \cdot m \cdot \left|\mathscr{G}_m^{g(m)}\right| \ . \tag{3.51}$$

Lemma 89. *Let $\epsilon > 0$, $p \in (0,1)$ constant and $g(n) = o\left(\frac{n}{\log(n)}\right)$ a fixed function. Then there exists $\alpha > 0$ and $n_0 \in \mathbb{N}$ such that for all $n \in \mathcal{S}_\epsilon$ and $n \geqslant n_0$ the following holds*

$$\mathbb{P}(G \in \mathscr{G}_n^{g(n)} \text{ has } \geqslant \alpha n \text{ pendant vertices}) \geqslant p \ . \tag{3.52}$$

Proof. Fix $g(n) = o\left(\frac{n}{\log(n)}\right)$, $n \in \mathcal{S}_\epsilon$ and $p \in (0,1)$. Let $\alpha = \frac{(1-p)}{2c_\epsilon}$ and denote by $\tilde{p}(n) = \mathbb{P}(G \in \mathscr{G}_n^{g(n)}$ has $\geqslant \alpha n$ leaves). To prove the lemma it is sufficient to show that $\tilde{p} \geqslant p$.

We now prove the lemma by constructing from each graph with few pendant vertices many graphs in $\mathscr{G}_{n+1}^{g(n+1)}$ with little double counting and as a result finding a bound on $\tilde{p}(n)$.

Let $G \in \mathscr{G}_n^{g(n)}$ and such that G has fewer than αn pendant vertices. There are exactly $(1 - \tilde{p}(n)) \cdot \left|\mathscr{G}_n^{g(n)}\right|$ such graphs. To construct a graph $G' \in \mathscr{G}_{n+1}^{g(n+1)}$ from G, we first pick one of the vertices in $\{1, \cdots, n+1\}$, x say. There are $n+1$ choices for this. We now put the graph G on the vertex set $[n+1] \setminus \{x\}$ in such a way that there is an increasing bijection from the vertex set of G to $[n+1] \setminus \{x\}$. The vertex x is now added to G as a pendant vertex incident to some vertex $y \in [n+1] \setminus \{x\}$. There are n choices for where to attach x to as a pendant vertex.

How often is each graph $G' \in \mathscr{G}_{n+1}^{g(n+1)}$ constructed? To get back to G from G', we just need to delete the vertex x and then move the vertex set from $[n+1] \setminus \{x\}$ to $[n]$ using

an order preserving bijection. How many choices for x are there? There are fewer than $\lceil \alpha n \rceil$ pendant vertices in G, that is at most $\lceil \alpha n \rceil$ pendant vertices in G'. Since x is a pendant vertex, there are hence at most $\lceil \alpha n \rceil$ possible choices for the vertex x. Hence, each graph G' is constructed at most αn times. We thus have

$$\left|\mathscr{G}_{n+1}^{g(n+1)}\right| \geqslant \left|\mathscr{G}_{n}^{g(n)}\right| \cdot (1-\tilde{p}(n)) \cdot \frac{n^2}{\alpha n} \ . \tag{3.53}$$

Since $n \in \mathcal{S}_\epsilon$ we know that

$$\left|\mathscr{G}_{n+1}^{g(n+1)}\right| \leqslant c_\epsilon \cdot n \cdot \left|\mathscr{G}_{n}^{g(n)}\right| \ . \tag{3.54}$$

Combining equation (3.53) and equation (3.54) we obtain

$$\left|\mathscr{G}_{n}^{g(n)}\right| \cdot (1-\tilde{p}(n)) \cdot \frac{n^2}{\lceil \alpha n \rceil} \leqslant \left|\mathscr{G}_{n+1}^{g(n+1)}\right| \leqslant c_\epsilon \cdot n \cdot \left|\mathscr{G}_{n}^{g(n)}\right| \ , \tag{3.55}$$

which implies that $1-\tilde{p}(n)$ must satisfy

$$\alpha \geqslant \frac{1-\tilde{p}(n)}{c_\epsilon} \ . \tag{3.56}$$

It follows that $\tilde{p}(n) \geqslant 1 - \alpha c_\epsilon = 1 - \frac{1-p}{2c_\epsilon} c_\epsilon \geqslant p$ is as required to prove the lemma. ■

We have thus far established that as long as $g(n) = O\left(\frac{n}{\log(n)}\right)$, the appearance of pendant vertices is still relatively frequent. We now use this to show that as a consequence the maximum degree is small relatively often. This result will then allow us to bound the size of $\mathscr{G}_n^{g(n)}$ above for most n as long as $g(n) = o\left(\frac{n}{\log^3(n)}\right)$. We denote in the following by Δ the maximum degree of a graph. Fix $\epsilon > 0$ for the remainder of this section. Then by Lemma 89 with $p = 1/2$ there exists an $\alpha > 0$ such that for all $n \in \mathcal{S}_\epsilon$ and n sufficiently large we have that

$$\mathbb{P}(G \in \mathscr{G}_n^{g(n)} \text{ has } \geqslant \alpha n \text{ pendant vertices}) \geqslant \frac{1}{2} \ . \tag{3.57}$$

We define the following sets on graphs in $\mathscr{G}_n^{g(n)}$ where we choose α is as in Lemma 89 for $p = 1/2$ and $c := 8/\alpha$ and $b = \lfloor c \ln(n) \rfloor$

PVI := set that G has at least αn pendant vertices

PVII := set that each vertex in G is adjacent to at most $2\ln(n)/\ln\ln(n)$ pendant vertices

MD := set that $\Delta(G) \leqslant b$.

We then have that for a graph G chosen uniformly at random from $\mathscr{G}_n^{g(n)}$ where $n \in \mathcal{S}_\epsilon$ and $g(n) = o\left(\frac{n}{\log(n)}\right)$

$$\mathbb{P}(R_n \in \mathbf{PVI}) \geqslant \frac{1}{2} \cdot (1 + o(1)) \ . \tag{3.58}$$

The following lemma provides a bound for the probability of $R_n \in \mathbf{PVII}$.

Lemma 90. *In $G \in \mathscr{G}_n^{g(n)}$ with high probability each vertex is adjacent to at most $2\ln(n)/\ln\ln(n)$ pendant vertices.*

Proof. The same proof as the proof of Lemma 2.2 in [43] applies. ■

From Lemma 90 it follows that

$$\mathbb{P}(R_n \in \mathbf{PVII}) \geqslant 1 - o(1) \ . \tag{3.59}$$

We now aim to find a bound on the probability of $R_n \in \mathbf{MD}$ occurring.

Lemma 91. *We have that for all $n \in \mathcal{S}_\epsilon$ and n sufficiently large*

$$\mathbb{P}(R_n \in \mathbf{MD}) \geqslant \frac{1}{2} \cdot (1 + o(1)) \ . \tag{3.60}$$

Proof. Let $\epsilon > 0$ and $n \in \mathcal{S}_\epsilon$. Let n_0 be such that for all $n \in \mathcal{S}_\epsilon$ and $n \geqslant n_0$ we have that $\mathbb{P}(R_n \in \mathbf{PVI}) \geqslant 1/2$. Now assume further that $n \geqslant n_0$. Note that by the same proof as the proof of Theorem 1.2 in [43] we know that

$$\mathbb{P}(R_n \in (\mathbf{MD}^C \cap \mathbf{PVI} \cap \mathbf{PVII})) = o(1) \tag{3.61}$$

and we know from equations (3.58) and (3.59) that

$$\mathbb{P}(R_n \in (\mathbf{PVI} \cap \mathbf{PVII})) \geqslant \frac{1}{2} \cdot (1 + o(1)) \ . \tag{3.62}$$

Thus

$$\begin{aligned} \mathbb{P}(R_n \in \mathbf{MD}^C) &\leqslant \mathbb{P}(R_n \in (\mathbf{PVI} \cap \mathbf{PVII})^C) + \mathbb{P}(R_n \in (\mathbf{MD}^C \cap \mathbf{PVI} \cap \mathbf{PVII})) \\ &\leqslant \frac{1}{2} \cdot (1 + o(1)) + o(1) \\ &= \frac{1}{2} \cdot (1 + o(1)) \ , \end{aligned}$$

as required. ■

This implies that at least $1/4$ of all graphs in $\mathscr{G}_n^{g(n)}$, when $n \in \mathcal{S}_\epsilon$ and n is sufficiently large, have maximum degree at most $c\ln(n)$. This allows us to bound the size of these graph classes as follows:

Lemma 92. *Let $\epsilon > 0$ and $g(n) = o\left(\frac{n}{\log^3(n)}\right)$ be fixed. Then as $n \to \infty$ for $n \in \mathcal{S}_\epsilon$*

$$\left|\mathscr{G}_n^{g(n)}\right| \leqslant \gamma_{\mathscr{P}}^n \cdot n! \cdot (1 + o(1))^n \ , \tag{3.63}$$

where $\gamma_{\mathscr{P}}$ is the planar graph growth constant.

Proof. Let $\epsilon > 0$ and $g(n) = o\left(\frac{n}{\log^3(n)}\right)$ be fixed. Further, assume that $n \in \mathcal{S}_\epsilon$ and that n is sufficiently large. Then we have that at least $1/4$ of all graphs in $\mathscr{G}_n^{g(n)}$ have maximum degree at most $c\ln(n)$. Recall that any graph with maximum degree at most Δ has an edge planarising set of size at most $c'\sqrt{ng(n)\Delta}$ for some constant c' [30, Theorem 4]. So let $G \in \mathscr{G}_n^{g(n)}$ with maximum degree at most $c\ln(n)$ and define $C = c' \cdot \sqrt{c}$. Then there exists a set of at most $C \cdot \sqrt{ng(n)\ln(n)}$ edges whose deletion leaves a planar graph G'. How often is each graph G' constructed? There are at most

$$\sum_{i=0}^{C\cdot\sqrt{ng(n)\ln(n)}} \binom{n^2}{i} \leqslant C \cdot \sqrt{ng(n)\ln(n)} \cdot n^{2C\cdot\sqrt{ng(n)\ln(n)}}$$

choices for which at most $C \cdot \sqrt{ng(n)\ln(n)}$ edges to add to G' to obtain G. Hence each graph G' is constructed at most $n^{2C\cdot\sqrt{ng(n)\ln(n)}}$ times. Since at least $1/4$ of all graphs in $\mathscr{G}_n^{g(n)}$ have maximum degree at most $c\ln(n)$ we have that

$$\begin{aligned}\left|\mathscr{G}_n^{g(n)}\right| &\leqslant 4 \cdot C \cdot \sqrt{ng(n)\ln(n)} \cdot n^{2C\cdot\sqrt{ng(n)\ln(n)}} \cdot |\mathscr{P}_n| \\ &\leqslant n^{2C\cdot\sqrt{ng(n)\ln(n)}} \cdot \gamma_l^n \cdot n! \cdot (1+o(1))^n \ ,\end{aligned}$$

which implies since $g(n) = o\left(\frac{n}{\log^3(n)}\right)$ that

$$\left|\mathscr{G}_n^{g(n)}\right| \leqslant \gamma_{\mathscr{P}}^n \cdot n! \cdot (1+o(1))^n \ , \tag{3.64}$$

as required. ■

We have thus found a bound on the size of $\left|\mathscr{G}_n^{g(n)}\right|$ for at least a constant density of all $n \in \mathbb{N}$ since $|\mathcal{S}_\epsilon|$ satisfies that for all $N \in \mathbb{N}$

$$|\mathcal{S}_\epsilon \cap [N]| \geqslant (1-\epsilon)N \ . \tag{3.65}$$

Using this, we can prove an upper bound on $\left|\mathscr{G}_n^{g(n)}\right|$ for all $n \in \mathbb{N}$ as follows:

Lemma 93. *Let $\epsilon > 0$ and $g(n) = o\left(\frac{n}{\log^3(n)}\right)$. Then as $n \to \infty$ (without any restriction)*

$$\left|\mathscr{G}_n^{g(n)}\right| \leqslant n! \cdot \gamma_{\mathscr{P}}^{(1+\epsilon)n} \cdot (1+o(1))^n \ . \tag{3.66}$$

Proof. Let $0 < \epsilon < 1$ and $g(n) = o\left(\frac{n}{\log^3(n)}\right)$ be fixed. As $n \to \infty$ with $n \in \mathcal{S}_{\epsilon/2}$ we know by Lemma 92 that

$$\left|\mathscr{G}_n^{g(n)}\right| \leqslant \gamma_{\mathscr{P}}^n \cdot n! \cdot (1+o(1))^n \ . \tag{3.67}$$

All that is left to show is that this is also satisfied for all $n \notin \mathcal{S}_{\epsilon/2}$. To do so, suppose $n \notin \mathcal{S}_{\epsilon/2}$. Then since $\left|\mathcal{S}_{\epsilon/2}\right|$ satisfies that for all $N \in \mathbb{N}$

$$\left|\mathcal{S}_{\epsilon/2} \cap [N]\right| \geqslant (1-\epsilon)N \ , \tag{3.68}$$

there exists an $m \in \mathbb{N}$ such that

$$n < m \leqslant (1+\epsilon)n \tag{3.69}$$

and

$$m \in \mathcal{S}_{\epsilon/2} \ . \tag{3.70}$$

Furthermore, note that by Lemma 76

$$\left|\mathscr{G}_{n+1}^{g(n+1)}\right| \geqslant (n+1)\left|\mathscr{G}_{n}^{g(n)}\right| \ . \tag{3.71}$$

From this, it follows that

$$\begin{aligned}\left|\mathscr{G}_{n}^{g(n)}\right| &\leqslant \frac{1}{(m-1)\cdots n}\left|\mathscr{G}_{m}^{g(m)}\right| \\ &\leqslant \frac{1}{(m-1)\cdots n} m! \cdot \gamma_{\mathscr{P}}^{m}(1+o(1))^{m} \\ &\leqslant n! \cdot \gamma_{\mathscr{P}}^{(1+\epsilon)n}(1+o(1))^{n}\end{aligned}$$

as $m \leqslant (1+\epsilon)n$ as required. ■

We are now in a position to prove our main result: the existence of a growth constant for all genus $g(n) = o\left(\frac{n}{\log^3(n)}\right)$.

Theorem 94. *For any $g(n) = o\left(\frac{n}{\log^3(n)}\right)$, the graph class $\mathscr{G}_n^{g(n)}$ has growth constant equal to the planar graph growth constant $\gamma_{\mathscr{P}}$, equivalently*

$$\left|\mathscr{G}_{n}^{g(n)}\right| \leqslant n! \cdot \gamma_{\mathscr{P}}^{n} \cdot (1+o(1))^{n} \ . \tag{3.72}$$

Proof. By Lemma 93, we have that

$$\limsup_{n\to\infty}\left(\frac{\left|\mathscr{G}_{n}^{g(n)}\right|}{n!}\right)^{1/n} \leqslant \gamma_{\mathscr{P}} \ , \tag{3.73}$$

and since also $\mathscr{P}_n \subseteq \mathscr{G}_n^{g(n)}$ and

$$\left(\frac{|\mathscr{P}_n|}{n!}\right)^{1/n} \to \gamma_{\mathscr{P}}$$

as n tends to infinity, we have that

$$\liminf_{n\to\infty}\left(\frac{\left|\mathscr{G}_{n}^{g(n)}\right|}{n!}\right)^{1/n} \geqslant \gamma_{\mathscr{P}} \ . \tag{3.74}$$

Putting equations (3.73) and (3.74) together, we obtain that

$$\left(\frac{\left|\mathscr{G}_{n}^{g(n)}\right|}{n!}\right)^{1/n} \to \gamma_{\mathscr{P}}$$

as n tends to infinity as required for $\mathscr{G}_n^{g(n)}$ to have growth constant $\gamma_{\mathscr{P}}$. ■

3.3.2.4 Alternative Proof of Growth Constant up to Genus $g(n) = o\left(\frac{n}{\log^{3+\epsilon}(n)}\right)$

In this section we give an alternative proof to show that the classes of graphs on n vertices and embeddable in the orientable or non-orientable surface of Euler genus at most any given $g(n) = O\left(\frac{n}{\log^{3+\epsilon}(n)}\right)$ for some $\epsilon > 0$ have a growth constant and this growth constant is equal to the planar graph growth constant. The proof in this section has a very different flavour to the proof in the last section and is split into two parts. We first prove in Section 3.3.2.4 the existence of a growth constant equal to the planar graph growth constant for graphs embeddable in the orientable or non-orientable surface of Euler genus at most $g(n)$, for any given $g(n) = o\left(\frac{n}{\log^4(n)}\right)$. This is done by repeatedly splitting the vertices on short non-contractible, non-surface-separating cycles until we have a planar graph. In Section 3.3.2.4, we then obtain the growth constant $\gamma_{\mathscr{P}}$ for all classes of graphs embeddable in the orientable surface up to any given Euler genus $g(n) = O\left(\frac{n}{\log^{3+\epsilon}(n)}\right)$ for some $\epsilon > 0$. We prove this by relating graphs embeddable in an orientable surface of "large" Euler genus to graphs embeddable in an orientable surface of, relative to the number of vertices, "smaller" Euler genus by adding a planar graph on "many" vertices to the graph. We then use the fact that a graph chosen uniformly at random from $\mathscr{G}_m^{g(m)}$ has with high probability maximum degree $O(\log(m))$ as long as a growth constant exists [43] and that for every graph G on n vertices and of Euler genus g and of maximum degree Δ, there exists by Theorem 75 a planarising edge set of size at most $O\left(\sqrt{n \cdot g \cdot \Delta}\right)$.

Growth Constant $\gamma_{\mathscr{P}}$ up to Euler Genus $o\left(\frac{n}{\log^4(n)}\right)$

We define $k(n,h) = \left\lfloor c_9 \cdot \sqrt{\frac{n}{h+1}} \cdot \log(h+1) \right\rfloor$ for some constant $c_9 > 0$ to be defined later, which will be the number of vertices we add when cutting along a non-separating cycle in the graph and splitting the vertices along said cycle in two. From Theorem 73 we have the following Lemma.

Lemma 95. *Given a constant $c_8 > 0$ and any triangulated orientable combinatorial surface without boundary (S,G), possibly with multi-edges, with Euler genus h, where $1 \leqslant h \leqslant c_8 \cdot n$, and n vertices, there exists a constant $c_9 > 0$ such that some non-separating closed curve along edges in (S,G) has length at most $c_9 \cdot \sqrt{\frac{n}{h+1}} \cdot \log(h+1)$.*

The same lemma holds for non-orientable surfaces for a different constant c_9. Note that Lemma 95 implies the following lemma, since every simple graph on a surface can be extended to a pseudograph triangulation of the same surface by adding edges.

Lemma 96. *Given a constant $c_8 > 0$ and any non-trivial orientable surface $\mathbf{S}$ of Euler genus $h \leqslant c_8 \cdot n$ and any graph $G \in \mathscr{G}_n^h$ cellularly embedded in $\mathbf{S}$, there exists a constant*

$c_9 > 0$ *such that there exists a non-separating cycle* C *in* $\mathbf{S}$ *which meets the graph* G *in at most* $k(n,h) = \left\lfloor c_9 \cdot \sqrt{\frac{n}{h+1}} \cdot \log(h+1) \right\rfloor$ *vertices.*

Note again that this also holds in the non-orientable case for a different constant c_9 and hence the following proof also extends to the non-orientable case. For simplicity of notation in the following proof, we define $K = k(2n,h)$ and note that for all $m \leqslant 2n$ we have that $k(m,h) \leqslant K$. The first key step to proving the existence of a growth constant is the following Lemma.

Lemma 97. *Let* h *be a positive integer,* n *a positive integer, assume* $h \leqslant c_8 \cdot n$ *for some constant* $c_8 > 0$*, define* $K = k(2n,h) = \left\lfloor c_9 \cdot \sqrt{\frac{2n}{h+1}} \cdot \log(h+1) \right\rfloor$ *and let* $m \leqslant 2n$*. Given a graph* $G \in \mathscr{G}^h_{m+K}$ *and a vector* $x \in [m]^K$*, let* $\psi(G,x)$ *denote the (multi-) graph obtained by starting with* G *and identifying vertices* x_j *and* $m+j$ *for each* $j = 1, \ldots, K$*. Then, for each graph* $G \in \mathscr{G}^h_m$*, where* $m \leqslant 2n$*, there exists a graph* $\tilde{G} \in \mathscr{G}^{h-1}_{m+K}$ *and a list* $x \in [m]^K$ *of* K *distinct vertices such that* $\psi(\tilde{G}, x) = G$*; and thus*

$$\left|\mathscr{G}^h_m\right| \leqslant \left|\mathscr{G}^{h-1}_{m+K}\right| \cdot m^K \ . \tag{3.75}$$

Proof. This can be shown as follows (following the proof in [42]): Let $G \in \mathscr{G}^h_m$ and let G be cellularly embedded in an orientable surface $\mathbf{S}$ of Euler genus at most $h \leqslant c_8 \cdot n$ for some constant $c_8 > 0$. By Lemma 96 there is a constant $c_9 > 0$ such that there exists a non-separating cycle C in $\mathbf{S}$ meeting G in k' vertices where $0 \leqslant k' \leqslant k(m,h) \leqslant K$.

List the vertices along C as $v_1, \ldots, v_{k'}$ and form a graph G' by cutting the surface along C splitting v_i into two vertices v_i and $m+i$ with edges incident to the original vertex v_i set in the natural way incident to either the new v_i or $m+i$ for each i (if $k' = 0$ then G' is just G).

Since C is non-surface-separating the newly constructed graph G' satisfies $G' \in \mathscr{G}^{h-1}_{m+k'}$. By adding isolated vertices if necessary, we can construct $\tilde{G} \in \mathscr{G}^{h-1}_{m+K}$ together with a list of exactly K distinct vertices in G such that $\psi(\tilde{G}, x) = G$. The result follows. ■

Proposition 98. *For any given function* $g(n) = o\left(\frac{n}{\log^4(n)}\right)$*, the classes of graphs* $\mathscr{G}^{g(n)}_n$ *and* $\mathscr{H}^{g(n)}_n$ *have a growth constant, which is equal to the planar graph growth constant.*

Proof. By Lemma 97, we know that for every $m \leqslant 2n$

$$|\mathscr{G}^{g(n)}_m| \leqslant |\mathscr{G}^{g(n)-1}_{m+k(2n,g(n))}| \cdot m^{k(2n,g(n))} = \left|\mathscr{G}^{g(n)-1}_{m+K}\right| \cdot m^K \ .$$

We will repeatedly cut along short non-separating cycles in the graph and split the vertices along the cycle. We hope, that by the time we have achieved a planar graph, we will

have added at most n new vertices, so can bound the number of vertices present in each step above by $2n$. Thus, for every $0 \leqslant m \leqslant 2n$ we have that

$$\begin{aligned}\frac{|\mathscr{G}_m^{g(n)}|}{m!} &\leqslant \frac{|\mathscr{G}_{m+k(2n,g(n))}^{g(n)-1}|}{(m+k(2n,g(n)))!} \cdot m^{k(2n,g(n))} \cdot (m+k(2n,g(n)))^{k(2n,g(n))} \\ &\leqslant \frac{|\mathscr{G}_{m+k(2n,g(n))}^{g(n)-1}|}{(m+k(2n,g(n)))!} \cdot (m+k(2n,g(n)))^{2\cdot k(2n,g(n))} \ .\end{aligned}$$

For clarity, we will from now on omit the dependence on n in the genus, so write g instead of $g(n)$. To simplify notation, we now define for $n \in \mathbb{N}$ and $g(n) \in \mathbb{N}$

$$f(n,g(n)) := \frac{\left|\mathscr{G}_n^{g(n)}\right|}{n!}$$

and for all $g(n) \in \mathbb{N}$

$$f(0,g) = 1 \ .$$

Thus the above implies for $m \leqslant 2n$

$$f(m,g) \leqslant f(m+k(2n,g),g-1) \cdot (m+k(2n,g))^{2\cdot k(2n,g)} \ . \tag{3.76}$$

We now repeatedly apply the above splitting step. We start off by splitting G along a non-separating cycle meeting at most $k(2n,g)$ vertices. In the process of doing so, we add $k(2n,g)$ vertices. We then repeat this procedure for the resulting graph and so on until we have either reached a planar graph or the number of vertices in the graph has become greater than $2n$ (which we never do in our construction, as we will show later on). In each step we add $k(2n,p)$ vertices as long as the number of vertices in the graph that we split is at most $2n$ and where p is the Euler genus of the graph that we split. Thus the total number of vertices we add is at most (assuming we have not violated the upper bound on the number of vertices)

$$\begin{aligned}\eta &= k(2n,g) + k(2n,g-2) + \ldots + k(2n,2) \\ &\leqslant 2 \cdot c_{11} \cdot \sqrt{n \cdot g} \cdot \log(n) \ ,\end{aligned}$$

for some constant $c_{11} > 0$. Since for $g(n) = o\left(\frac{n}{\log^4(n)}\right)$, we have $n + \eta < 2n$ as long as n is large enough implying that we can do this until we get to a planar graph. We further know from [34] that for all m and some constant $c_{10} > 0$ the following holds

$$f(m,0) \leqslant c_{10} \cdot \gamma_{\mathscr{P}}^m \ .$$

Note that after at most $g(n)$ (in the orientable case it should be exactly $g(n)/2$) splitting steps the graph G has been turned into a planar graph on $n+\eta$ vertices, where,

as long as $g(n) = o\left(\frac{n}{\log^4(n)}\right)$, we have $\eta = o\left(\frac{n}{\log(n)}\right)$. Thus by equation (3.76) we have

$$\begin{aligned} f(n, g(n)) &\leqslant (n + K(g))^{2K(g)} \cdot f(n + K(g), g(n) - 1) \\ &\leqslant (n + \eta)^{2K(g)} \cdot f(n + K(g), g(n) - 1) \\ &\leqslant (n + \eta)^{2\eta} \cdot f(n + \eta, 0) \\ &\leqslant (n + \eta)^{2\eta} \cdot c_{10} \cdot \gamma_{\mathscr{P}}^{n+\eta} \ . \end{aligned}$$

It follows that we have, since $\eta = o\left(\frac{n}{\log(n)}\right)$, for $g(n) = o\left(n/\log^4(n)\right)$

$$\begin{aligned} \left(\frac{\left|\mathscr{G}_n^{g(n)}\right|}{n!}\right)^{\frac{1}{n}} &= (f(n, g(n)))^{\frac{1}{n}} \\ &\leqslant \left((n + \eta)^{2\eta} \cdot c_{10} \cdot \gamma_{\mathscr{P}}^{n+\eta}\right)^{\frac{1}{n}} \\ &= \gamma_{\mathscr{P}} \cdot \left((n + \eta)^2 \cdot \gamma_{\mathscr{P}}\right)^{\frac{\eta}{n}} \cdot c_{10}^{\frac{1}{n}} \\ &= \gamma_{\mathscr{P}} \cdot (1 + o(1)) \ . \end{aligned} \tag{3.77}$$

Since also $\left(\frac{|\mathscr{P}_n|}{n!}\right)^{\frac{1}{n}} \to \gamma_{\mathscr{P}}$ and $\mathscr{P}_n \subset \mathscr{G}_n^{g(n)}$, we see that

$$\left(\frac{\left|\mathscr{G}_n^{g(n)}\right|}{n!}\right)^{\frac{1}{n}} \to \gamma_{\mathscr{P}}$$

as $n \to \infty$ as long as $g(n) = o\left(\frac{n}{\log^4(n)}\right)$. As a result, $\mathscr{G}_n^{g(n)}$ has growth constant $\gamma_{\mathscr{P}}$ as long as $g(n) = o\left(\frac{n}{\log^4(n)}\right)$. ■

As mentioned before, the above proof also works in the non-orientable case for a different constant c_9. This implies that for $g(n) = o\left(\frac{n}{\log^4(n)}\right)$, the class of graphs $\mathscr{H}_n^{g(n)}$ has growth constant equal to the planar graph growth constant $\gamma_{\mathscr{P}}$. Note again that by using Observation 66 we could also have deduced the orientable case from the non-orientable case.

Growth Constant up to Euler Genus $O\left(\frac{n}{\log^{3+\epsilon}(n)}\right)$

We have thus far established, that all graph classes $\mathscr{G}_n^{g(n)}$ of any given $g(n) = o\left(\frac{n}{\log^4(n)}\right)$ have a growth constant, which is equal to the planar graph growth constant $\gamma_{\mathscr{P}}$. We now wish to extend this up to any given Euler genus $g(n) = O\left(\frac{n}{\log^{3+\epsilon}(n)}\right)$ for any fixed $\epsilon > 0$.

Proposition 99. *For any given function $g(n) = O\left(\frac{n}{\log^{3+\epsilon}(n)}\right)$ for some $\epsilon > 0$, the class of graphs $\mathscr{G}_n^{g(n)}$ and $\mathscr{H}_n^{g(n)}$ have a growth constant, which is equal to the planar graph growth constant.*

We will prove this by using properties of graph classes that have a growth constant and relating graphs embeddable in an orientable surface of "large" Euler genus to graphs embeddable in an orientable surface of, relative to the number of vertices, "smaller" Euler genus by adding a planar graph on "many" vertices to the graph.

To do so, we take a graph G in $\mathscr{G}_n^{g(n)}$ and construct from it many graphs G' in $\mathscr{G}_{n\cdot f(n)}^{g(n)}$ with little double counting, where $f(n)$ is a function satisfying both of the following:

- $f(n) > 2$ for all $n \in \mathbb{N}$
- $f(n) = o(n)$.

Once we have picked a labeled graph G, we choose n labels from $[n \cdot f(n)]$ to relabel the vertices of G using an order preserving bijection from $[n]$ to the subset of size n of $[n \cdot f(n)]$. In what follows, these n vertices will be called "special" vertices. On the remaining $n \cdot f(n) - n$ vertices, we put any connected planar graph P in $\mathscr{C}_{n\cdot f(n)-n}$, where $\mathscr{C}_k$ denotes the class of connected planar graphs on k vertices (we will actually take $f(n) \approx \log^{\delta}(n)$ for very small δ).

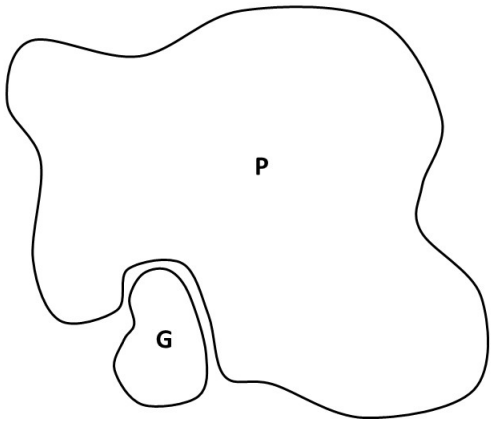


Figure 3.11: The graph G'

There are $\binom{n\cdot f(n)}{n}$ ways of choosing the special vertices and $\left|\mathscr{C}_{n\cdot f(n)-n}\right|$ possible connected planar graphs to be put on the remaining vertices. We thus construct from the graph G

$$\binom{n \cdot f(n)}{n} \left|\mathscr{C}_{n\cdot f(n)-n}\right|$$

graphs $G' \in \mathscr{G}_{n\cdot f(n)}^{g(n)}$.
By [34] we know that for some constant $c_{12} > 0$

$$\left|\mathscr{C}_{n\cdot f(n)-n}\right| \sim c_{12} \cdot (n \cdot f(n) - n)^{-\frac{7}{2}} \cdot \gamma_{\mathscr{P}}^{n\cdot f(n)-n} \cdot (n \cdot f(n) - n)! \; . \tag{3.78}$$

This implies that for n large enough the number of constructions from a given graph $G \in \mathscr{G}_n^{g(n)}$ is at least

$$\frac{1}{2} \cdot \frac{(n \cdot f(n))!}{n!} \cdot c_{12} \cdot (n \cdot f(n) - n)^{-\frac{7}{2}} \cdot \gamma_{\mathscr{P}}^{n \cdot f(n) - n} \ .$$

How often is each graph G' constructed? Since $n \cdot f(n) > 2n$ we have that there exists only one component of size $n \cdot f(n) - n$, the planar connected component. The vertices not in this component are the "special" vertices. Each graph G' is thus constructed exactly once. Therefore

$$\left|\mathscr{G}_{n \cdot f(n)}^{g(n)}\right| \geqslant \binom{n \cdot f(n)}{n} \cdot \left|\mathscr{C}_{n \cdot f(n) - n}\right| \cdot \left|\mathscr{G}_n^{g(n)}\right| \ .$$

Rearranging,

$$\left|\mathscr{G}_n^{g(n)}\right| \leqslant \frac{\left|\mathscr{G}_{n \cdot f(n)}^{g(n)}\right|}{\binom{n \cdot f(n)}{n} \cdot \left|\mathscr{C}_{n \cdot f(n) - n}\right|} \ . \tag{3.79}$$

How large is the graph class $\mathscr{G}_{n \cdot f(n)}^{g(n)}$? Suppose that $g(n)$ is small enough to give us that $\mathscr{G}_{n \cdot f(n)}^{g(n)}$ has a growth constant, not necessarily equal to the planar graph growth constant. We will find precise values for $f(n)$ that work for our purpose later on. From [43] we know that the existence of a growth constant implies that with high probability the maximum degree of a graph chosen uniformly at random from the graph class $\mathscr{G}_{n \cdot f(n)}^{g(n)}$ is $O(\log(n))$. As a result a graph chosen uniformly at random from $\mathscr{G}_{n \cdot f(n)}^{g(n)}$ has by [30] with high probability a planarizing set of edges of size

$$O\left(\sqrt{n \cdot f(n) \cdot g(n) \cdot \log(n)}\right) \ . \tag{3.80}$$

This implies that there exists a constant $c_{13} > 0$ such that

$$\left|\mathscr{G}_{n \cdot f(n)}^{g(n)}\right| \leqslant n^{c_{13} \cdot \sqrt{n \cdot f(n) \cdot g(n) \cdot \log(n)}} \cdot \left|\mathscr{P}_{n \cdot f(n)}\right| \ . \tag{3.81}$$

We know from [34], that for some constant $c_2 > 0$

$$|\mathscr{P}_n| \sim c_2 \cdot n^{-\frac{7}{2}} \cdot \gamma_{\mathscr{P}}^n \cdot n! \ ,$$

implying that as long as $\mathscr{G}_{n \cdot f(n)}^{g(n)}$ has a growth constant we can bound its size above as follows:

$$\left|\mathscr{G}_{n \cdot f(n)}^{g(n)}\right| \leqslant n^{c_{13} \cdot \sqrt{n \cdot f(n) \cdot g(n) \cdot \log(n)}} \cdot c_2 \cdot (n \cdot f(n))^{-\frac{7}{2}} \cdot \gamma_{\mathscr{P}}^{n \cdot f(n)} \cdot (n \cdot f(n))! \ . \tag{3.82}$$

Putting this into equation (3.79), we get for n large enough and some constant $c_{14} > 0$

$$\begin{aligned}\left(\frac{\left|\mathscr{G}_n^{g(n)}\right|}{n!}\right)^{\frac{1}{n}} &\leqslant \left(\frac{n^{c_{13}\cdot\sqrt{n\cdot f(n)\cdot g(n)\cdot\log(n)}}\cdot c_2\cdot(n\cdot f(n))^{-\frac{7}{2}}\cdot\gamma_{\mathscr{P}}^{n\cdot f(n)}\cdot(n\cdot f(n))!}{n!\cdot\frac{(n\cdot f(n))!}{n!}\cdot c_{12}\cdot(n\cdot f(n)-n)^{-\frac{7}{2}}\cdot\gamma_{\mathscr{P}}^{n\cdot f(n)-n}}\right)^{\frac{1}{n}} \\ &\leqslant \left(\gamma_{\mathscr{P}}^n\cdot n^{c_{13}\cdot\sqrt{n\cdot f(n)\cdot g(n)\cdot\log(n)}}\cdot c_{14}\right)^{\frac{1}{n}} \\ &= \gamma_{\mathscr{P}}\cdot n^{c_{13}\cdot\sqrt{\frac{f(n)\cdot g(n)\cdot\log(n)}{n}}}\cdot c_{14}^{\frac{1}{n}}\ ,\end{aligned}$$

which tends to $\gamma_{\mathscr{P}}$ as $n\to\infty$ as long as the exponent of n is $o\left(1/\log(n)\right)$. We also know that

$$\left(\frac{|\mathscr{P}_n|}{n!}\right)^{\frac{1}{n}}\to\gamma_{\mathscr{P}}$$

and $\mathscr{P}_n\subset\mathscr{G}_n^{g(n)}$, meaning that as long as the exponent of n is $o\left(1/\log(n)\right)$ and $\mathscr{G}_{n\cdot f(n)}^{g(n)}$ has a growth constant, the class of graphs $\mathscr{G}_n^{g(n)}$ also has a growth constant equal to the planar graph growth constant $\gamma_{\mathscr{P}}$.
Equivalently, we need the following two conditions to hold for $\mathscr{G}_n^{g(n)}$ to have a growth constant equal to the planar graph growth constant:

$$\mathscr{G}_{n\cdot f(n)}^{g(n)}\text{ has a growth constant} \tag{3.83}$$

$$f(n)\cdot g(n)\cdot\log(n\cdot f(n))=o\left(\frac{n}{\log^2(n)}\right)\ . \tag{3.84}$$

So far our proof works for all functions $f(n)$ and $g(n)$, as long as they satisfy conditions (3.83) and (3.84). We now prove that for every $\epsilon>0$ there exists a function $f(n)$ such that for $g(n)=\frac{n}{\log^{3+\epsilon}(n)}$ conditions (3.83) and (3.84) are satisfied. This implies that for every $\epsilon>0$, the class of graphs $\mathscr{G}_n^{g(n)}$ for any given $g(n)=O\left(\frac{n}{\log^{3+\epsilon}(n)}\right)$ has a growth constant.

To prove the existence of such a function $f(n)$, we now define

$$A=\{x\in\mathbb{R}^+\mid\mathscr{G}_n^{\frac{n}{\log^x(n)}}\text{ has growth constant }\gamma_{\mathscr{P}}\}\ .$$

We have proven in Section (3.3.1) that the class of graphs embeddable in the orientable surface of any given Euler genus $g(n)=\omega\left(\frac{n}{\log(n)}\right)$ does not have a growth constant, implying $A\subseteq[1,\infty)$. Furthermore, in the last part of this section, we showed that $\mathscr{G}_n^{n/\log^4(n)}$ does have a growth constant, so $4\in A$ and so A is non-empty. Now let $\alpha=\inf\{x\mid x\in A\}$. We aim to show that $\alpha\leqslant 3$.

In the previous part of this section, it was shown that the class of graphs embeddable in the orientable surface of any given Euler genus $g(n) = o\left(\frac{n}{\log^4(n)}\right)$ has a growth constant equal to the planar graph growth constant $\gamma_{\mathscr{P}}$, that is $\alpha \leqslant 4$. Now suppose we have shown that $\beta \in A$. Then letting $\delta > 0$, $\eta > 0$ and $f(n) = \log^{\delta}(n)$ and $g(n) = \frac{n}{\log^{\eta}(n)}$, we have

$$g(n) = \frac{n}{\log^{\eta}(n)} = \frac{n \cdot \log^{\delta}(n)}{\log^{\eta+\delta}(n)} \leqslant \frac{2n \cdot f(n)}{\log^{\eta+\delta}(n \cdot f(n))} < \frac{n \cdot f(n)}{\log^{\beta}(n \cdot f(n))} \ .$$

So as long as $\eta + \delta > \beta$, which implies $\mathscr{G}_{n \cdot f(n)}^{g(n)}$ has growth constant $\gamma_{\mathscr{P}}$, Condition (3.83) is satisfied. Furthermore, as long as

$$f(n) \cdot g(n) \cdot \log(n \cdot f(n)) \leqslant 2 \cdot \log^{\delta}(n) \cdot \frac{n}{\log^{\eta}(n)} \cdot \log(n) = \frac{2n}{\log^{\eta-\delta-1}(n)} = o\left(\frac{n}{\log^2(n)}\right) \ ,$$

Condition (3.84) is also satisfied. Hence Conditions (3.83) and (3.84) being satisfied are equivalent to the following two conditions being satisfied

$$\eta + \delta > \beta \tag{3.85}$$

$$\eta - \delta > 3 \ . \tag{3.86}$$

Now suppose for a contradiction that $\alpha > 3$. Then there exists a $\epsilon > 0$ such that $\alpha - 4\epsilon > 3$. We define

$$\beta = \alpha + \epsilon$$
$$\eta = \alpha - \epsilon$$
$$\delta = 3\epsilon \ .$$

Since α is defined to be the infimum of $A = \{x \in \mathbb{R}^+ \mid \mathscr{G}_n^{\frac{n}{\log^x(n)}}$ has growth constant $\gamma_{\mathscr{P}}\}$, we have that $\alpha + \epsilon \in A$, that is β is well defined. Furthermore, $\eta + \delta = \alpha - \epsilon + 3\epsilon > \beta$, by definition of β. Thus Condition (3.85) is satisfied. Note that by the definition of ϵ, we have that $\eta - \delta = \alpha - \epsilon - 2\epsilon > 3$, implying that condition (3.86) is also satisfied. Hence by the above proof, we have that $\eta = \alpha - \epsilon \in A$, contradicting α being the infimum.

Thus all graphs classes $\mathscr{G}_n^{g(n)}$, for any given $g(n) = O\left(\frac{n}{\log^{3+\epsilon}(n)}\right)$ for some $\epsilon > 0$, have growth constant equal to the planar graph growth constant $\gamma_{\mathscr{P}}$. Note that this proof also works in the non-orientable case for a different constant for the length of a shortest non-separating cycle.

3.4 Properties of Graph Classes

In this section we give results on common properties of graphs chosen uniformly at random from the class of graphs $\mathscr{G}_n^{g(n)}$ or $\mathscr{H}_n^{g(n)}$ for different functions $g(n)$. So far we have managed to find some results, however few precise results are known thus far and we hope to continue our work on the topics in this section in the future. We do further make some conjectures on how we expect the properties to behave for larger $g(n)$.

3.4.1 Number of Edges and Number of Faces

Observation 69 gives us a bound on the maximum number of edges any graph of genus at most h can have. It follows directly that no graph in $\mathscr{G}_n^{g(n)}$ and $\mathscr{H}_n^{g(n)}$ can have more than $3(n + g(n) - 2)$ edges. Further, Observation 72 gives a bound on the minimum number of edges a connected graph of genus at least g can have. We know that most graphs in the graph classes $\mathscr{G}_n^{g(n)}$ and $\mathscr{H}_n^{g(n)}$ have genus exactly $g(n)$ or $g(n) - 1$. Is it possible to deduce a lower bound on the number of edges for most graphs in the graph classes $\mathscr{G}_n^{g(n)}$ or $\mathscr{H}_n^{g(n)}$ from this? The answer is yes. In the following we prove three such bounds. While the bounds get more precise in each result, the range of $g(n)$ for which they hold gets smaller. To prove the first most general bound we require the following Lemma from [41].

Lemma 100. *[41, Lemma 2.3] Let $\mathscr{A}$ be a bridge-addable class of graphs, and let $R_n \in_U \mathscr{A}_n$. Then for each $\epsilon > 0$*

$$\mathbb{P}[\mathit{frag}(R_n) > 2/\epsilon] < \epsilon \ .$$

Furthermore, we need to know that for all $g(n)$, even when $g(n) \sim \alpha n^2$ for some constant $\alpha > 0$, a graph chosen uniformly at random from $\mathscr{G}_n^{g(n)}$ or $\mathscr{H}_n^{g(n)}$ has with large probability genus close to $g(n)$. We prove this in the following lemma:

Lemma 101. *Let $\delta > 0$ and $0 < \eta < 1/12$. Then there is an integer k such that the following holds. Let $g(n) \geqslant 0$ be a genus function, let $\bar{g}(n) = \min\{g(n), \eta n^2\}$ and let $R_n \in_U \mathscr{A}_n^{\bar{g}(n)}$, where $\mathscr{A}_n^{\bar{g}(n)} = \mathscr{G}_n^{\bar{g}(n)}$ or $\mathscr{A}_n^{\bar{g}(n)} = \mathscr{H}_n^{\bar{g}(n)}$. Then*

$$\mathbb{P}(R_n \notin \mathscr{A}_n^{\bar{g}(n)-k}) \geqslant 1 - \delta$$

for sufficiently large n.

Proof. Together with equation (3.14), it then follows that for any fixed $j \in \mathbb{N}$

$$\begin{aligned}
\left|\mathscr{A}_n^{g(n)+2-j}\right| &\geqslant \left|\mathscr{A}_n^{\bar{g}(n)+2-j}\right| \\
&\geqslant \frac{\binom{n}{2} - 3n - 3(\bar{g}(n) - j)}{3n + 3(\bar{g}(n) - j)} \cdot \left|\mathscr{A}_n^{\bar{g}(n)-j}\right| \\
&\geqslant \alpha(1 + O(1/n)) \left|\mathscr{A}_n^{\bar{g}(n)-j}\right|
\end{aligned}$$

where $\alpha = \frac{1-6\eta}{6\eta}$.

Noting that $\alpha > 1$ since $0 < \eta < 1/12$, we see that for some constant $j > 0$ and n sufficiently large

$$\left|\mathscr{A}_n^{\bar{g}(n)}\right| \geqslant \left(\frac{1}{\delta}\right) \left|\mathscr{A}_n^{\bar{g}(n)-j}\right| \ . \tag{3.87}$$

Now let $k = j$ to get the required result. ■

We can now prove the following Proposition:

Proposition 102. *Let $0 < \epsilon < 1$ and $0 < \eta < 1/12$. Then there is an integer k such that the following holds. Let $g(n) \geqslant 0$ be a genus function, let $\bar{g}(n) = \min\{g(n), \eta n^2\}$, and let $R_n \in_U \mathscr{A}_n^{\bar{g}(n)}$, where $\mathscr{A}_n^{\bar{g}(n)} = \mathscr{G}_n^{\bar{g}(n)}$ or $\mathscr{A}_n^{\bar{g}(n)} = \mathscr{H}_n^{\bar{g}(n)}$. Then*

$$\mathbb{P}(e(R_n) \geqslant \bar{g}(n) + n - k) \geqslant 1 - \epsilon$$

for n sufficiently large.

Proof. We start by applying Lemma 101 with $\delta = \epsilon/2$ to obtain a constant k_1 such that

$$\mathbb{P}(R_n \in \mathscr{G}_n^{\bar{g}(n)-k}) < \epsilon/2 \ . \tag{3.88}$$

Note that by Lemma 100 and since $\mathscr{G}_n^{k(n)}$ is bridge-addable for any function $k(n)$, there exists a constant $k_2 = \lfloor 1/\epsilon \rfloor$ such that

$$\mathbb{P}(\mathrm{frag}(R_n) > k_2) < \epsilon/2 \ . \tag{3.89}$$

We thus have that with probability at least $1 - \epsilon$ both $R_n \notin \mathscr{G}_n^{\bar{g}(n)-k_1}$ and $\mathrm{frag}(R_n) < k_2$ hold. In this case, to make R_n connected, we need to add at most $k_2 - 1$ edges between the fragment and the big component containing at least $n - k_2 + 1$ vertices. A connected graph of genus at least $\bar{g} - k_1$ must have at least $\bar{g} - k_1 + n - 1$ edges by Observation 72. The connected graph we have constructed from R_n must hence also have at least this many edges. This implies, since we have added no more than $k_2 - 1$ edges, that R_n has at least

$$\begin{aligned} &\bar{g}(n) - k_1 + n - 1 - k_2 + 1 \\ &= \bar{g}(n) + n - (k_1 + k_2) \end{aligned}$$

edges. Since this is true with probability at least $1 - \epsilon$ and by choosing $k = k_1 + k_2$, the result follows. ■

We now find a bound on the minimum number of faces in a typical graph in $\mathscr{G}_n^{g(n)}$ and $\mathscr{H}_n^{g(n)}$. This will help us, in combination with Euler's formula (see Theorem 67), to prove that for genus functions $g(n) = o(n^2)$, a graph chosen uniformly at random from $\mathscr{G}_n^{g(n)}$ or $\mathscr{H}_n^{g(n)}$ has with high probability at least $n + g(n)$ edges.

Lemma 103. *Let $\epsilon > 0$. Then there is a constant $c > 0$ such that the following holds. Let $g(n)$ be a genus function and let $R_n \in_U \mathscr{A}_n^{g(n)}$, where $\mathscr{A}_n^{g(n)} = \mathscr{G}_n^{g(n)}$ or $\mathscr{A}_n^{g(n)} = \mathscr{H}_n^{g(n)}$. Then the probability that any embedding of R_n in the orientable or non-orientable surface of Euler genus $g(n)$ respectively has less than $c \cdot \frac{n^2}{n+g(n)}$ faces is smaller than ϵ.*

This lemma implies that as long as $g(n) = O(n)$ with high probability we have a linear number of faces.

Proof. With foresight, let us set $c = \epsilon/14$. We prove the above by a double counting argument. For this, we define the following subclass of $\mathscr{A}_n$

$$\mathscr{B}_n = \left\{ G \in \mathscr{A}_n \mid G \text{ has an embedding with } \leqslant c \cdot \frac{n^2}{n+g(n)} \text{ faces} \right\} ,$$

where by embedding here we mean embedding in an appropriate surface and where c is a constant we define later. Let $G \in \mathscr{B}_n$ have f faces say. We now allocate each vertex of G to one of the faces it is incident to. By adding an edge between vertices that have been allocated to the same face (as long as the edge is not already present in G), we create a graph G' with one more face than G. Note that G' is also certainly contained in the class $\mathscr{A}_n$. Note that we avoid counting any potential edges twice by assigning each vertex to a unique face. We denote by n_i the number of vertices we have allocated to face f_i. There are thus exactly

$$\begin{aligned} \sum_{i=1}^{f} \binom{n_i}{2} &= \left(\sum_{i=1}^{f} \frac{n_i^2}{2} \right) - \left(\sum_{i=1}^{f} \frac{n_i}{2} \right) \\ &\geqslant \left(\frac{1}{2} \cdot \sum_{i=1}^{f} \left(\frac{n}{f} \right)^2 \right) - \frac{n}{2} \\ &= \frac{n^2}{2f} - \frac{n}{2} \end{aligned}$$

edges between vertices that have been allocated to the same face. Further, there are at most $3(g(n)+n)$ edges already present in the graph G. Hence, from G we construct at least $\frac{n^2}{2f} - \frac{n}{2} - 3(g(n)+n)$ graphs $G' \in \mathscr{A}_n$.

How often is each graph constructed? There are at most $3(g(n)+n)$ edges in G'. Guess the new edge to return to G, there are at most $3(g(n)+n)$ choices for this. Hence, each graph G' is constructed at most $3(g(n)+n)$ times. Thus

$$|\mathscr{B}_n| \leqslant \frac{3(g(n)+n)}{\frac{n^2}{2f} - \frac{n}{2} - 3(g(n)+n)} \cdot |\mathscr{A}_n| , \tag{3.90}$$

assuming that the denominator is greater than zero. From this follows directly that

$$\begin{aligned}
\mathbb{P}(G \in \mathscr{B}_n) &= \frac{|\mathscr{B}_n|}{|\mathscr{A}_n|} \\
&\leqslant \frac{3(g(n)+n)}{\frac{n^2}{2f} - \frac{n}{2} - 3(g(n)+n)} \\
&\leqslant \frac{3(g(n)+n)}{\frac{g(n)+n}{2c} - \frac{n}{2} - 3(g(n)+n)} \\
&\leqslant \frac{1}{\frac{1}{6c} - \frac{n}{6(g(n)+n)} - 1} \\
&\leqslant \frac{1}{\frac{1}{6c} - \frac{7}{6}} \\
&= \frac{6c}{1-7c} \,,
\end{aligned}$$

but $1 - 7c > \frac{1}{2}$, so $\frac{6c}{1-7c} < 12c < \epsilon$, which completes the proof. ■

This now allows us to prove the following theorem concerning the number of edges in a graph chosen uniformly at random from $\mathscr{G}_n^{g(n)}$ or $\mathscr{H}_n^{g(n)}$, which is the second lower bound on the number of edges with high probability that we find. While the bound is tighter than the first, note that unlike the first it does not hold in the case $g(n) = \Theta(n^2)$.

Theorem 104. *Let $g(n) \geqslant 0$ and $g(n) = o(n^2)$ be a genus function and $R_n \in_U \mathscr{A}_n$, where $\mathscr{A}_n = \mathscr{G}_n^{g(n)}$ or $\mathscr{A}_n = \mathscr{H}_n^{g(n)}$. Then $R_n \in_U \mathscr{A}_n$ has with high probability at least $g(n) + n$ edges.*

Proof. Let $\epsilon > 0$ and $R_n \in \mathscr{A}_n$. Note that by [46, Theorem 2.2] and since $\mathscr{G}_n^{g(n)}$ is bridge-addable for any function $g(n)$ the number of components of R_n, denoted by $\kappa(R_n)$, satisfies $\mathbb{E}[\kappa(R_n)] < 2$. Thus, applying Markov's inequality, we have that for any function $m(n)$, such that $m(n) \to \infty$ and for some n sufficiently large

$$\mathbb{P}(\kappa(R_n) \geqslant m(n)) < \frac{2}{m(n)} < \frac{\epsilon}{3} \,. \tag{3.91}$$

It follows that $R_n \in \mathscr{A}_n$ has with probability at least $1 - \epsilon/3$ less than $m(n)$ components.

Note that by Lemmas 78 and 79, R_n has with high probability a (connected or disconnected) cellular embedding in a surface of Euler genus $g(n)$ or $g(n) - 1$. By Euler's formula for disconnected graphs (see Theorem 68) we hence know that for n sufficiently large with probability at least $1 - \epsilon/3$ either

$$\begin{aligned}
2 - g(n) &= n - e - \kappa + 1 + f \quad \text{or} \\
2 - g(n) + 1 &= n - e - \kappa + 1 + f \,,
\end{aligned} \tag{3.92}$$

where e is the number of edges in R_n, f is the number of faces and κ is the number of components in the cellular embedding. By Lemma 103, with probability at least $1-\epsilon/3$, the embedding of R_n has at least $c \cdot \frac{n^2}{n+g(n)}$ faces. Substituting this into (3.92) and noting that by the above, choosing $m(n) = c \cdot \frac{n^2}{n+g(n)} - 4$, R_n has with probability at most $\epsilon/3$ more than $\frac{n^2}{n+g(n)} - 2$ components. Hence, with probability at least $1-\epsilon$ for n sufficiently large, we have either

$$e = n + g - 2 - c \cdot \frac{n^2}{n+g(n)} + 4 + c \cdot \frac{n^2}{n+g(n)} = n + g + 2 \quad \text{or}$$
$$e = n + g - 3 - c \cdot \frac{n^2}{n+g(n)} + 4 + c \cdot \frac{n^2}{n+g(n)} = n + g + 1 \ .$$

So as long as

$$-1 \leqslant c \cdot \frac{n^2}{n+g(n)} - c \cdot \frac{n^2}{n+g(n)} \ ,$$

which clearly always holds, we have that $e \geqslant n + g(n)$ with probability at least $1-\epsilon$, as required. ■

Another way of bounding the number of edges present in a graph chosen uniformly at random from $\mathscr{G}_n^{g(n)}$ or $\mathscr{H}_n^{g(n)}$ is by considering the number of edges an edge maximal graph in the class can have and noting that by Theorem 2 of [28] the expected number of edges is at least half that many edges.

Theorem 105. *Let $R_n \in \mathscr{A}_n$, where $\mathscr{A}_n = \mathscr{G}_n^{g(n)}$ or $\mathscr{A}_n = \mathscr{H}_n^{g(n)}$, and $g(n) = o(n)$. Then the expected number of edges in R_n is at least $\frac{3}{2}n \cdot (1 + o(1))$.*

Proof. Let $R_n \in \mathscr{A}_n$, where $\mathscr{A}_n = \mathscr{G}_n^{g(n)}$ or $\mathscr{A}_n = \mathscr{H}_n^{g(n)}$. By [47], we have that any edge-maximal embedding of a graph embeddable in $\mathbf{S}_{g(n)/2}$ or $\mathbf{N}_{g(n)}$ is at most $O(g(n))$ edges short of a triangulation. Applying Theorem 2 of [28] then implies that a graph chosen uniformly at random from $\mathscr{A}_n$ has expected number of edges at least

$$\begin{aligned}
&\frac{1}{2} \cdot (\text{number of edges of a triangulation } - O(g(n))) \\
&\quad = \frac{1}{2} \cdot (3n + 3g(n) - 3 - O(g(n))) \\
&\quad \geqslant \frac{3}{2} \cdot n \cdot (1 + o(1))
\end{aligned}$$

edges for n sufficiently large, as required. ■

3.4.2 Number of Pendant Vertices

We know in the planar case, and in fact whenever a growth constant exists, that a random graph from the class has with failure probability $e^{-\Omega(n)}$ a linear number of pendant vertices

by Theorem 59. On the other hand we know that the Erdős–Rényi random graph with edge probability 1/2 has with high probability no vertices of degree one since, if we consider $G(n,p)$, then as n tends to infinity the minimum degree is $\sim np > 1$. This means that a graph chosen uniformly at random from $\mathscr{G}_n^{k(n)}$, where $k(n)$ is the maximum genus of a graph on n vertices, has with high probability no pendant vertices since $R_n \in \mathscr{G}_n^{g(n)}$ behaves the same as $G(n, \frac{1}{2})$. Note that the following results should also hold for pendant appearances of any connected planar graph H. We focus here on pendant vertices, since the number of pendant vertices heavily influences other properties like the maximum degree as can be seen in the following sections.

So far, we have only managed to establish bounds for the number of pendant vertices for genus $g(n) = O\left(\frac{n}{\log(n)}\right)$ and for genus $g(n) = \omega(n)$ and no results for genus between $\omega(n/\log(n))$ and $\Theta(n)$. We are currently divided about whether these upper bounds are tight or whether there is actually a sharp threshold at either $n/\log(n)$ or n, where the number of pendant vertices changes from being linear with high probability to being at most constant. We now present our current results on the number of pendant vertices in the different regions of $g(n)$.

3.4.2.1 Class of Graphs of Euler Genus $g(n) = O\left(\frac{n}{\log(n)}\right)$

We know that for all classes of graphs that are small, the subclass of graphs with less than a linear number of pendant vertices has strictly bigger radius of convergence of the exponential generating function than the original graph class. This is stated formally in the following theorem.

Theorem 106. *[44] Let $\mathscr{A}_n$, where $\mathscr{A}_n = \mathscr{G}_n^{g(n)}$ the class of graphs of at most any given orientable Euler genus $g(n) = O\left(\frac{n}{\log(n)}\right)$, or $\mathscr{A}_n = \mathscr{H}_n^{g(n)}$ the class of graphs of at most any given non-orientable Euler genus $g(n) = O\left(\frac{n}{\log(n)}\right)$, then the exponential generating function has radius of convergence $0 < \rho < \infty$. Then $\alpha = \frac{\rho}{2e}$ is such that the following holds: Let $\mathscr{B}$ be the class of graphs in $\mathscr{A}_n$ with less than αn pendant vertices then*

$$\rho(\mathscr{B}) > \rho \ .$$

We thus expect there to be linearly many pendant vertices in this case as long as the size of the graph class grows sufficiently smooth. We have not managed to prove the smoothness of $\mathscr{G}_n^{g(n)}$ or $\mathscr{H}_n^{g(n)}$ for $g(n) = \omega\left(\frac{n}{\log^3(n)}\right)$ yet but expect it to hold as long as the genus function $g(n)$ also behaves sufficiently smooth in some sense.

3.4.2.2 Class of Graphs of Euler Genus $o\left(\frac{n}{\log^3(n)}\right)$

Since these classes all have a growth constant, which is equal to the planar graph growth constant $\gamma_{\mathscr{P}}$, these classes all have with high probability linearly many pendant appearances of any fixed planar graph by Theorem 59. We wish to improve the constant for which this holds when H is just a single vertex. The proof is the same as in [44, Theorem 2.13], however when just considering pendant vertices rather than pendant attachments of general planar graphs we can slightly improve on the constant α.

Theorem 107. *[45] Let the class $\mathscr{A}_n$, where $\mathscr{A}_n = \mathscr{G}_n^{g(n)}$ or $\mathscr{A}_n = \mathscr{H}_n^{g(n)}$, of graphs have a growth constant γ. Then $\alpha = \frac{\rho}{2e} > 0$, where $\rho = \gamma^{-1}$, is such that for $R_n \in_U \mathscr{G}$*

$$\mathbb{P}\left(R_n \text{ has } \leqslant \alpha n \text{ leaves}\right) = e^{-\Omega(n)} \ .$$

3.4.2.3 Class of Graphs of Euler Genus $g(n) \gg n$

Proposition 108. *For any given $g(n) \gg n$ such that $g(n) = o(n^2)$, a graph chosen uniformly at random from $\mathscr{A}_n$, where $\mathscr{A}_n = \mathscr{G}_n^{g(n)}$ or $\mathscr{A}_n = \mathscr{H}_n^{g(n)}$, has with high probability at most $\frac{2n^2}{g(n)}$ pendant vertices.*

Proof. By Theorem 104 we know that with high probability a graph chosen uniformly at random from $\mathscr{G}_n^{g(n)}$ has at least $g(n)$ edges as long as $g(n) = o(n^2)$. We define the following two classes,

$$\begin{aligned}\mathscr{A}_n &:= \mathscr{G}_n^{g(n)} \\ \mathscr{B}_n &:= \{G \in \mathscr{G}_n^{g(n)} \mid G \text{ has } > 2 \cdot \frac{n^2}{g(n)} \text{ pendant vertices and } \geqslant g(n) \text{ edges}\} \ .\end{aligned}$$

We have that $G \in \mathscr{G}_n^{g(n)}$ for $g(n)$ has with high probability at least $g(n)$ edges by Theorem 104. It hence suffices to show that with high probability $G \notin \mathscr{B}_n$. The idea of the proof is to show that from the graphs in $\mathscr{B}_n$ we can build many graphs in $\mathscr{A}_n$ with little double counting, so we can not have started with many graphs in $\mathscr{B}_n$.

Let $G \in \mathscr{B}_n$ and pick $k = \frac{n^2}{g(n)}$ of its more than $2 \cdot \frac{n^2}{g(n)}$ pendant vertices and call them $v_1, \ldots, v_k$. First delete the edge incident to these vertices and then add the pendant vertices in the middle of an edge each. Since there are at least $g(n)$ edges in G, we have at least

$$\binom{2k}{k} \cdot g(n)^k$$

choices of where to put the pendant vertices, since two of the vertices v_i, v_j can also be placed on the same edge. Let the resulting graph be G'. Then $G' \in \mathscr{G}_n^{g(n)}$ and G' has fewer pendant vertices than G.

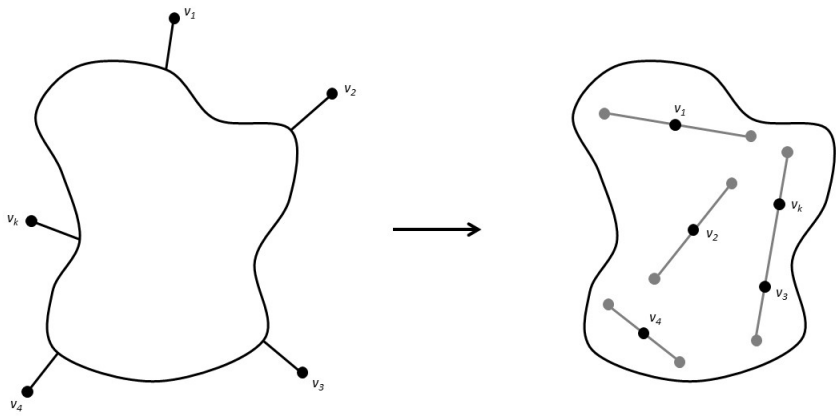


Figure 3.12: $G \in \mathscr{B}_n$ to $G' \in \mathscr{A}_n$

How often is each such graph G' constructed? First we guess the k vertices $v_1, \ldots, v_k$, there are at most $\binom{n}{k}$ choices. Now we guess the k vertices that the vertices $v_1, \ldots, v_k$ were originally attached to. There are at most n^k choices for this. So every graph G' is constructed at most

$$\binom{n}{k} \cdot n^k \leqslant \left(\frac{n \cdot e}{k}\right)^k \cdot n^k = (g(n) \cdot e)^k$$

times. It follows that

$$|\mathscr{A}_n| \geqslant |\mathscr{B}_n| \cdot \left(\frac{g(n)}{g(n) \cdot e}\right)^k \cdot \binom{2k}{k} .$$

So the proportion of graphs G in $\mathscr{B}_n$ is, when $g(n) \geqslant c \cdot n^\alpha$ for some constant $c > 0$

$$\begin{aligned} \frac{|\mathscr{B}_n|}{|\mathscr{A}_n|} &\leqslant \left(\frac{g(n) \cdot e}{g(n)}\right)^k \cdot \binom{2k}{k}^{-1} \\ &\leqslant e^k \cdot \binom{2k}{k}^{-1} \\ &= o(1) \quad \text{since } \binom{2k}{k} = (4 + o(1))^k \text{ and } 4 > e \ . \end{aligned} \tag{3.93}$$

Consequently, for genus $g(n)$ a graph chosen uniformly at random from $\mathscr{G}_n^{g(n)}$ does with high probability not have more than $\frac{2n^2}{g(n)}$ pendant vertices as required. ■

3.4.3 Connectivity

We know that a random planar graph has probability of being connected bounded away from both zero and one. On the other hand, the Erdős–Rényi random graph with edge probability $1/2$ and equivalently also a graph chosen uniformly at random from $\mathscr{G}_n^{k(n)}$, where $k(n)$ is the maximum orientable Euler genus of a graph on n vertices, is with high

probability connected. We expect the threshold of this property to be at either $n/\log(n)$ or n. Thus far we can only prove that the probability of connectedness is bounded away from one as long as $g(n) = o(n/\log^3(n))$ and is tending to one as n tends to infinity as long as $g(n) = \omega(n)$.

Note further, that we know that that for all bridge-addable classes a non-vanishing fraction of the graphs in the class are connected as Colin McDiarmid, Angelika Steger and Dominic JA Welsh have shown in Theorem 2.2 of [45]. Since any of the classes $\mathscr{G}_n^{g(n)}$ and $\mathscr{H}_n^{g(n)}$ are bridge-addable for any $g(n)$, we have the following theorem.

Theorem 109. *[46, Theorem 2.2] Let $\mathscr{A}$ be any finite non-empty bridge-addable class of graphs, that is for example $\mathscr{A} = \mathscr{G}_n^{g(n)}$ or $\mathscr{A} = \mathscr{H}_n^{g(n)}$ for some given non-negative, function $g(n)$ and let $R \in_U \mathscr{A}$, that is R is a graph sampled uniformly at random from $\mathscr{A}$. The random number $\kappa(R)$ of components of R is then stochastically dominated by $1+X$ where $X \sim Po(1)$, that is X has the Poisson distribution with mean 1. This further implies the following results*

- $\mathbb{P}(R \text{ is connected}) \geqslant \frac{1}{e}$
- $\mathbb{P}(\kappa(R) \geqslant k+1) \leqslant \frac{1}{k!}$
- $\mathbb{E}[\kappa(R)] \leqslant 2$.

3.4.3.1 Class of Graphs of Euler Genus $g(n) = o\left(\frac{n}{\log^3(n)}\right)$

However, when a growth constant exists, more can be said about the probability of connectedness. It turns out that for all classes of graphs with a growth constant, a graph chosen uniformly at random from this class of graphs has asymptotically a non-zero probability of not being connected.

Theorem 110. *[42, Theorem 2.5] Let $\mathscr{A}$ be any of $\mathscr{G}_n^{g(n)}$ or $\mathscr{H}_n^{g(n)}$ for a given $g(n)$ such that the class has a growth constant. Then for $R_n \in_U \mathscr{A}$ and for any fixed planar graph H,*

$$\liminf_{n\to\infty} \mathbb{P}(\mathit{Frag}(R_n) \cong H) > 0$$

and thus $\limsup_{n\to\infty} \mathbb{P}(R_n \text{is connected}) < 1$.

3.4.3.2 Class of Graphs of Euler Genus $O\left(\frac{n}{\log(n)}\right)$

While we can not yet make an absolute statement about all genus functions in this range, we can use the fact that the radius of convergence of the graph classes of these genus functions is positive to give us a partial result.

Proposition 111. *Let $g(n)$ be a non-decreasing function with $g(n) = O(n/\log(n))$. Write $\mathscr{A}_n$ for $\mathscr{G}_n^{g(n)}$ or $\mathscr{H}_n^{g(n)}$. Let $R_n \in_U \mathscr{A}_n$. Then*

$$\liminf \mathbb{P}(R_n \text{ is connected}) < 1 \ . \tag{3.94}$$

Proof. By Theorems 86 and 87, we know that the radius of convergence of $\mathscr{A}$ is strictly larger than zero. This then implies that

$$\limsup a_n^{1/n} < \infty \ ,$$

where $a_n = |\mathscr{A}_n|/n!$. From this it follows that also

$$\liminf a_n^{1/n} < \infty \ .$$

We can deduce from this result that further

$$\liminf \frac{a_n}{a_{n-1}} < \infty \ .$$

Now let $\liminf a_n/a_{n-1} < C < \infty$. There are infinitely many integers n such that $a_n/a_{n-1} \leqslant C$. Pick one such n. Then $|\mathscr{A}_n| \leqslant nC\,|\mathscr{A}_{n-1}|$. Suppose $\mathscr{C}_n$ is the set of connected graphs $G \in \mathscr{A}_n$. By bridge-addability and Theorem 109, we have

$$|\mathscr{C}_n| \geqslant |\mathscr{A}_n|/e \ .$$

For each graph $G \in \mathscr{C}$, the graph G' obtained by adding an isolated vertex is in $\mathscr{A}$, since we restricted $g(n)$ to be non-decreasing. Hence

$$\begin{aligned} |\mathscr{A}_n \setminus \mathscr{C}_n| &\geqslant n\,|\mathscr{C}_{n-1}| \\ &\geqslant \frac{n}{e}\,|\mathscr{A}_{n-1}| \\ &\geqslant \frac{n}{e}\left(\frac{1}{nC}\right)|\mathscr{A}_n| \end{aligned}$$

and thus $\mathbb{P}(R_n \text{ is connected}) \leqslant 1 - \left(\frac{1}{eC}\right)$, so $\liminf \mathbb{P}(R_n \text{ is connected}) \leqslant 1 - \frac{1}{eC}$, as required. ■

3.4.3.3 Class of Graphs of Euler Genus $g(n) = \omega(n)$

Theorem 112. *Let $\epsilon > 0$. Then there exist constants $n_0 > 0$ and $C > 0$ such that for each $n \geqslant n_0$ and each $h \geqslant Cn$, $R_n \in_U \mathscr{A}_n$, where $\mathscr{A}_n = \mathscr{G}_n^h$ or $\mathscr{A}_n = \mathscr{H}_n^h$, satisfies*

$$\mathbb{P}(R_n \text{ is connected}) \geqslant 1 - \epsilon \ .$$

It follows that if $g(n) \gg n$ and $R_n \in_U \mathscr{A}_n$, where $\mathscr{A}_n = \mathscr{G}_n^{g(n)}$ or $\mathscr{A}_n = \mathscr{H}_n^{g(n)}$, then R_n is connected with high probability.

Proof. Let k be the constant we obtain when applying Proposition 102 with $\epsilon/2$, so k is such that

$$\mathbb{P}(e(R_n) \geqslant \bar{g}(n) + n - k) \geqslant 1 - \epsilon/2$$

and from the proof of Proposition 102 and the choice of k in said proof, we also know that $\mathbb{P}(\text{frag}(R_n) > k) < \epsilon/2$.

We now wish to show that $\mathbb{P}(R_n \text{ is not connected}) \to 0$ as n tends to infinity. To do so we define

$$\begin{aligned} \mathscr{B}_n &= \mathscr{A}_n \\ \mathscr{C}_n &= \{G \in \mathscr{G}_n \mid 1 \leqslant \text{frag}(G) \leqslant k \cap e(G) \geqslant Cn + k\} \ . \end{aligned} \tag{3.95}$$

Since by Proposition 102, $R_n \in \mathscr{A}_n$ has with probability at least $1 - \epsilon/2$ at least

$$\bar{g}(n) + n - k \geqslant Cn + n - k \geqslant Cn + \binom{k}{2}$$

edges for n sufficiently large and fragment size strictly smaller than k, it suffices to show that

$$|\mathscr{C}_n| \leqslant \frac{\epsilon}{2} |\mathscr{B}_n| \ ,$$

in order to prove the theorem.

Define $\mathscr{C}_n^l$ to be the subclass of $\mathscr{C}_n$ containing all graphs whose fragment has size exactly l. Then

$$|\mathscr{C}_n| = \left|\mathscr{C}_n^1\right| + \ldots + \left|\mathscr{C}_n^k\right| \ .$$

Now let $G \in \mathscr{C}_n^l$. We wish to construct from G many graphs in $\mathscr{B}_n$ with little double counting. To construct a graph G' in $\mathscr{B}_n$ delete all the edges in $\text{Frag}(G)$ and add the vertices of $\text{Frag}(G)$ in the middle of any of the at least $g(n)$ edges of G. Call the resulting graph G'. There are at least $(Cn)^l$ choices for where to add the vertices, so we construct from one graph G at least $(Cn)^l$ graphs G'.

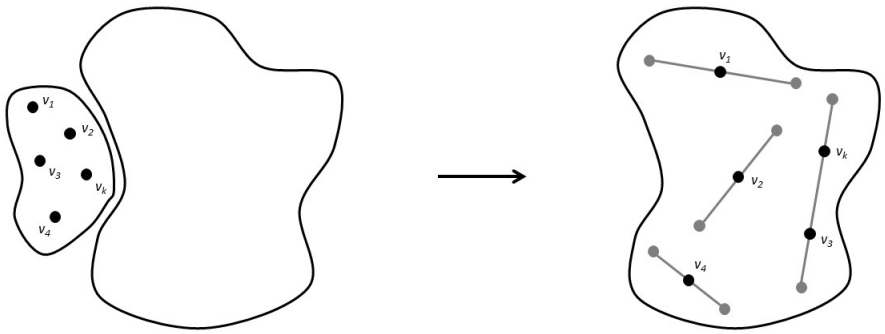


Figure 3.13: $G \in \mathscr{C}_n$ to $G' \in \mathscr{B}_n$

How often is each graph G' constructed? Each graph G' has n vertices and as a result there are at most $\binom{n}{l}$ choices of vertices in the fragment of G. Furthermore, there are at most $2^{\binom{l}{2}}$ choices for adding edges between the vertices in the fragment of G. This implies that

$$\begin{aligned}
|\mathscr{B}_n| &\geqslant \left|\mathscr{C}_n^l\right| \cdot \frac{(Cn)^l}{\binom{n}{l} \cdot 2^{\binom{l}{2}}} \\
&\geqslant \left|\mathscr{C}_n^l\right| \cdot \frac{(Cn)^l}{\left(\frac{n \cdot e}{l}\right)^l \cdot 2^{\frac{l(l-1)}{2}}} \\
&\geqslant \left|\mathscr{C}_n^l\right| \cdot \frac{l^l}{e^l \cdot 2^{\frac{l(l-1)}{2}}} \cdot \left(\frac{Cn}{n}\right)^l
\end{aligned} \tag{3.96}$$

and so

$$\begin{aligned}
\left|\mathscr{C}_n^l\right| &\leqslant |\mathscr{B}_n| \cdot \frac{e^l \cdot 2^{\frac{l(l-1)}{2}}}{l^l} \cdot \left(\frac{n}{Cn}\right)^l \\
&\leqslant |\mathscr{B}_n| \cdot \frac{e^l \cdot 2^{\frac{l(l-1)}{2}}}{l^l} \cdot C^{-l} \ .
\end{aligned} \tag{3.97}$$

It follows that

$$\begin{aligned}
|\mathscr{C}_n| &= \left|\mathscr{C}_n^1\right| + \ldots + \left|\mathscr{C}_n^k\right| \\
&\leqslant |\mathscr{B}_n| \cdot \left(\frac{e^1 \cdot 2^0}{1} \cdot C^{-1} + \ldots + \frac{e^k \cdot 2^{\frac{k^2}{2}}}{k^k} \cdot C^{-k}\right) \\
&\leqslant |\mathscr{B}_n| \cdot k \cdot e \cdot C^{-1} \\
&< \frac{\epsilon}{2} |\mathscr{B}_n| \ ,
\end{aligned} \tag{3.98}$$

for C sufficiently large, meaning $C > \frac{2ke}{\epsilon}$, and as we were required to show in order to prove the result. ■

3.4.4 Maximum Degree

Similarly to connectedness and the number of pendant vertices above, the maximum degree is known to behave very differently in the planar case to the uniformly at random case. In the planar case, we have that the maximum degree is with high probability of order $\Theta(\log(n))$, while in the Erdős–Rényi random graph with edge probability $1/2$ and hence also in $\mathscr{G}_n^{k(n)}$, where $k(n)$ is the maximum orientable Euler genus an n-vertex graph can have, the maximum degree is expected to be of order $\Theta(n)$. We again expect the transition between the two to happen either around $n/\log(n)$ or n. In terms of maximum degree, we have thus far only found some prelimiary results, however expect the maximum degree for genus $g(n) = \omega(n \log(n))$ to be roughly of order $\frac{g(n)}{n} \cdot \log(n)$.

We have already seen that in the case when $\mathscr{G}_n^{g(n)}$ or $\mathscr{H}_n^{g(n)}$ has a growth constant, the maximum degree behaves as in the planar case, meaning it is with high probability $\Theta(\log(n))$ by [43]. By Lemmas 78 and 79, a graph chosen uniformly at random from $\mathscr{G}_n^{g(n)}$ or $\mathscr{H}_n^{g(n)}$ has with high probability a cellular embedding on $\mathbf{S}_{g(n)/2}$ or $\mathbf{N}_{g(n)}$ respectively. Furthermore, by Theorem 104, a graph on n vertices chosen uniformly at random from $\mathscr{G}_n^{g(n)}$, for $g(n) = o(n^2)$, has with high probability at least $g(n) + n$ edges. The average degree of such a graph is thus with high probability at least $g(n)/n+1$, implying that the maximum degree is certainly at least this large. The same argument can be applied in the non-orientable case. As a result, the maximum degree of a graph chosen uniformly at random from $\mathscr{G}_n^{g(n)}$ or $\mathscr{H}_n^{g(n)}$ for $g(n) = \omega(n)$ is at least $g(n)/n+1$ with high probability.

3.5 Discussion of Threshold Behaviour

In Section 3.1.2 we have noted that classes of graphs embeddable on any fixed surface behave asymptotically as n tends to infinity exactly the same as planar graphs. On the other hand, a graph chosen uniformly at random from $\mathscr{G}_n^{k(n)}$, where $k(n)$ is the maximum orientable Euler genus a graph on n vertices can possess, behaves the same as an Erdős–Rényi random graph with edge probability $1/2$. The random planar and the Erdős–Rényi random graph behave very differently in terms of common properties a graph chosen uniformly at random from the class possesses. As an example, the random planar graph has with high probability linearly many pendant vertices by the Appearances Theorem (see Theorem 59), has with high probability maximum degree of order $\Theta(\log(n))$ by Theorem 60 and the probability of being connected is bounded away from one by Theorem 56. What is more, the class of planar graphs is small and in fact has a growth constant, the planar graph growth constant $\gamma_{\mathscr{P}}$. On the other hand, the Erdős–Rényi random graph has with high probability no pendant vertices, has with high proability maximum degree of order $\Theta(n)$ and is with high probability connected. Furthermore, the class of all graphs is over-large and certainly has no growth constant.

One of the main aims of this thesis and future work is to try to understand when the threshold of these properties occurs and whether the transition is sharp or gradual. From our work in this thesis, we can already make some statements about this threshold. As such, we have observed in Proposition 84 that for any given function $g(n) = \omega\left(\frac{n}{\log(n)}\right)$, the class of graphs $\mathscr{G}_n^{g(n)}$ and $\mathscr{H}_n^{g(n)}$ are over-large. On the other hand, by Theorems 86 and 87 the classes of graphs $\mathscr{G}_n^{g(n)}$ and $\mathscr{H}_n^{g(n)}$ are small for any given function $g(n) = O\left(\frac{n}{\log(n)}\right)$. This implies that the transition from small to over-large classes occurs at $\Theta(n/log(n))$. We furthermore expect the classes of graphs $\mathscr{G}_n^{g(n)}$ and $\mathscr{H}_n^{g(n)}$ to both have a growth

constant equal to the planar graph growth constant as long as $g(n) = o(n/\log(n))$. This seems a reasonable expectation as long as the classes grow sufficiently smoothly. One of the main issues that we have come across when trying to find bounds on the sizes of graph classes is that the classes $\mathscr{G}_n^{g(n)}$ and $\mathscr{H}_n^{g(n)}$ are not necessarily closed under removing a vertex if $g(n)$ is increasing.

In terms of the threshold of properties we can not make quite as precise statements yet. We have managed to show that there are with high probability no longer linearly many pendant vertices when $g(n) = \omega(n)$, maximum degree is certainly larger than $\log(n)$ as long as $g(n) \gg n\log(n)$ and for $g(n) = \omega(n)$ a graph chosen uniformly at random from the class is with high probability connected. We have however failed to make any statements about these properties for genus functions $g(n)$ that lie in the range $\omega(n/log(n)) = g(n) = o(n)$. This is since most of the results for large genus $g(n) = \omega(n)$ are deduced from a graph chosen uniformly at random having many, that is of order $\Theta(g(n))$, edges with high probability. However, as long as $g(n) = o(n)$, the number of edges is still not significantly larger than in the planar case and still of order $\Theta(n)$.

Since we expect all graph classes $\mathscr{G}_n^{g(n)}$ and $\mathscr{H}_n^{g(n)}$ for $g(n) = o(n/\log(n))$ to have a growth constant, we also expect the properties of these graph classes to be the same as in the planar case. Further, we expect the threshold of all properties to occur either at $\Theta(n/\log(n))$, where the threshold from small to over-large lies, or at $\Theta(n)$. The threshold at $\Theta(n)$ also appears a viable option since the classes are special in the sense that they are not necessarily closed under removing a vertex. If the closure under removing a vertex was given, some statements on common properties can be made from the size of a class. We are currently divided about the sharpness of the threshold of the appearance of pendant vertices and the maximum degree. It would be feasible to assume that the number of pendant vertices remains of order $\Theta(n)$ until genus $g(n) = \Theta(n)$ and then gradually decreases once we move past genus $g(n) = n$ to at most $\frac{2n^2}{g(n)}$, which is the upper bound we have observed in Proposition 108. It however also appears feasible to assume that there exists a sharp threshold for the number of pendant vertices and we move from having with high probability linearly many pendant vertices to having with high probaility at most a constant number of pendant vertices at either $g(n) = \Theta(n)$ or $g(n) = \Theta(n/\log(n))$.

For the maximum degree it appears reasonable to assume that it also remains of order $\log(n)$ until we move past genus $g(n) = \Theta(n)$, as the average degree of these graphs is still constant and since we do not expect the maximum degree to deviate from the average degree too far. Once we look at genus $g(n) \gg n\log(n)$, the maximum degree is certainly

with high probability larger than in the planar case as we have shown in Section 3.4.4. In this case, we expect the maximum degree to deviate from the average degree by at most a factor of $\log(n)$ as we can observe in the Erdős–Rényi random graph model for large average degree, this means we expect the maximum degree to be of order $\frac{g(n)}{n} \cdot \log(n)$ for genus $g(n) = \omega(n)$.

3.6 Unlabeled Graphs

We add a few results about unlabeled graphs here. Note that for unlabeled graphs we consider generating functions rather than exponential generating functions. The generating function $\widetilde{G}(x)$ of a class of unlabeled graphs $\widetilde{\mathscr{G}}$ is defined as follows:

$$\widetilde{G}(x) = \sum_{n \geqslant 0} \left|\widetilde{\mathscr{G}}_n\right| \cdot x^n \ .$$

So for the size of a class of unlabeled graphs $\widetilde{\mathscr{G}}$ we have the following definitions.

Definition. *A class of unlabeled graphs $\widetilde{\mathscr{G}}$ is **small** if the radius of convergence of its generating function is greater than zero, so if there exists a constant $c > 0$ such that there exists an $N \in \mathbb{N}$ such that the following holds*

$$\left|\widetilde{\mathscr{G}}_n\right|^{\frac{1}{n}} < c$$

for each $n \geqslant 0$.

Definition. *We say that a class of unlabeled graphs $\widetilde{\mathscr{G}}$ has **growth constant** $0 < \gamma < \infty$ if as n tends to infinity*

$$\left|\widetilde{\mathscr{G}}_n\right|^{\frac{1}{n}} \to \gamma \ .$$

Definition. *A class of unlabeled graphs $\widetilde{\mathscr{G}}_n$ is said to be **over-large** if the radius of convergence of its generating function is equal to zero, that is if for every constant $c > 0$ and for every $N \in \mathbb{N}$ there exists an $n \geqslant N$ such that*

$$\left|\widetilde{\mathscr{G}}_n\right|^{\frac{1}{n}} > c \ .$$

We define $\widetilde{\mathscr{G}}_n^{g(n)}$ to be the class of unlabeled graphs embeddable on an orientable surface of Euler genus at most $g(n)$, and similarly we define the class $\widetilde{\mathscr{H}}_n^{g(n)}$ to be the class of unlabeled graphs embeddable on a non-orientable surface of Euler genus at most $g(n)$. Using a supermultiplicativity argument, the existence of an unlabeled planar graph

growth constant $\widetilde{\gamma}_{\mathscr{P}}$ was shown in [28]. This means that for the class of unlabeled planar graphs $\widetilde{\mathscr{P}}$ it is known that

$$\left|\widetilde{\mathscr{P}}_n\right|^{\frac{1}{n}} \to \widetilde{\gamma}_{\mathscr{P}} \ ,$$

where $\gamma_{\mathscr{P}} < \widetilde{\gamma}_{\mathscr{P}} < 30.061$ as was shown in [25][45]. It was furthermore shown in [48] that for any fixed genus h the classes of graphs $\widetilde{\mathscr{G}}_n^h$ and $\widetilde{\mathscr{H}}_n^h$ have growth constant equal to the unlabeled planar graph growth constant, so

$$\begin{aligned} \left|\widetilde{\mathscr{G}}_n^h\right|^{\frac{1}{n}} &\to \widetilde{\gamma}_{\mathscr{P}} \\ \left|\widetilde{\mathscr{H}}_n^h\right|^{\frac{1}{n}} &\to \widetilde{\gamma}_{\mathscr{P}} \ . \end{aligned} \tag{3.99}$$

So far the exact value of $\widetilde{\gamma}_{\mathscr{P}}$ is however still unknown.

3.6.1 Lower Bounds

Note that every unlabeled n-vertex graph gives rise to at most $n!$ different labeled graphs, that is the number of labeled graphs divided by $n!$ gives a lower bound on the number of unlabeled graphs. So

$$\begin{aligned} \left|\widetilde{\mathscr{G}}_n^{g(n)}\right| &\geqslant \frac{\left|\mathscr{G}_n^{g(n)}\right|}{n!} \\ \left|\widetilde{\mathscr{H}}_n^{g(n)}\right| &\geqslant \frac{\left|\mathscr{H}_n^{g(n)}\right|}{n!} \ . \end{aligned} \tag{3.100}$$

Hence, we get directly from Section 2.1 the following lower bounds on the size of $\widetilde{\mathscr{G}}_n^{g(n)}$ and $\widetilde{\mathscr{H}}_n^{g(n)}$:

- always for $g(n) < \frac{\alpha n}{\log(n)}$ where $\alpha \leqslant 2\log\left(\frac{\widetilde{\gamma}_{\mathscr{P}}}{\gamma_{\mathscr{P}}}\right)$, where $\frac{\widetilde{\gamma}_{\mathscr{P}}}{\gamma_{\mathscr{P}}} \in (1, 1.1041)$, from the bound on $\widetilde{\mathscr{P}}(n)$ in [28]

$$\begin{aligned} \left|\widetilde{\mathscr{G}}_n^{g(n)}\right| &\geqslant \left|\widetilde{\mathscr{P}}_n\right| = \widetilde{\gamma}_{\mathscr{P}}^n \cdot (1 + o(1))^n \\ \left|\widetilde{\mathscr{H}}_n^{g(n)}\right| &\geqslant \left|\widetilde{\mathscr{P}}_n\right| = \widetilde{\gamma}_{\mathscr{P}}^n \cdot (1 + o(1))^n \ , \end{aligned} \tag{3.101}$$

- when $g(n) \geqslant \frac{\alpha n}{\log(n)}$ for $\alpha \geqslant 2\log\left(\frac{\widetilde{\gamma}_{\mathscr{P}}}{\gamma_{\mathscr{P}}}\right)$, where $\frac{\widetilde{\gamma}_{\mathscr{P}}}{\gamma_{\mathscr{P}}} \in (1, 1.1041)$, from equation (3.24)

$$\begin{aligned} \left|\widetilde{\mathscr{G}}_n^{g(n)}\right| &\geqslant \gamma_{\mathscr{P}}^n \cdot n^{\frac{g(n)}{2}} \cdot (1 + o(1))^n \\ \left|\widetilde{\mathscr{H}}_n^{g(n)}\right| &\geqslant \gamma_{\mathscr{P}}^n \cdot n^{\frac{g(n)}{2}} \cdot (1 + o(1))^n \ , \end{aligned} \tag{3.102}$$

- when there exists a $c_0 > 0$ such that $\omega\left(\frac{n}{\log(n)}\right) = g(n) \leqslant c_0 \cdot n$ holds then there exists a constant $c_1 > 0$ depending on c_0 such that, from equation (3.20)

$$\begin{aligned} \left|\widetilde{\mathscr{G}}_n^{g(n)}\right| &\geqslant c_1^n \cdot n^{g(n)} \\ \left|\widetilde{\mathscr{H}}_n^{g(n)}\right| &\geqslant c_1^n \cdot n^{g(n)} \ , \end{aligned} \tag{3.103}$$

- when $g(n) = \omega(n)$, from equations (3.19)

$$\begin{aligned} \left|\widetilde{\mathscr{G}}_n^{g(n)}\right| &\geqslant e^n \cdot \left(\frac{n^2 - 3n}{2(n + g(n))}\right)^{g(n)} \\ \left|\widetilde{\mathscr{H}}_n^{g(n)}\right| &\geqslant e^n \cdot \left(\frac{n^2 - 3n}{2(n + g(n))}\right)^{g(n)} \ . \end{aligned} \tag{3.104}$$

Note also that, as in the labelled case, the inequality $\left|\widetilde{\mathscr{H}}_n^{g(n)+1}\right| \geqslant \left|\widetilde{\mathscr{G}}_n^{g(n)}\right|$ holds with the same proof as in the orientable case. Hence, all of the bounds on $\left|\widetilde{\mathscr{H}}_n^{g(n)+1}\right|$ could equally just have been deduced from the orientable case. We thus have, as in the labeled case, that all classes of graphs of Euler genus any given $g(n) \gg \frac{n}{\log(n)}$ are over-large. Note furthermore that since the planar graph growth constant is strictly smaller than the unlabeled planar graph growth constant, the bounds for small genera are actually no improvements on the size of $\widetilde{\mathscr{G}}_n^{g(n)}$ and $\widetilde{\mathscr{H}}_n^{g(n)}$ compared to bounding it by the planar bound until $g(n) \geqslant \frac{\alpha n}{\log(n)}$ for $\alpha \geqslant 2\log\left(\frac{\tilde{\gamma}_{\mathscr{P}}}{\gamma_{\mathscr{P}}}\right)$, where $\frac{\tilde{\gamma}_{\mathscr{P}}}{\gamma_{\mathscr{P}}} \in (1, 1.1041)$.

3.6.2 Upper Bounds

We now try to find upper bounds on the size of $\widetilde{\mathscr{G}}_n^{g(n)}$ using the same method as in Section 3.3.2.1. This is done by counting the number of one-face embedded graphs on n vertices and then counting the number of ways edges can be added in the face. Unfortunately, we do not have a bound on the fractions of graph in $\widetilde{\mathscr{G}}_n^{g(n)}$ that are connected and cellularly embeddable in $\mathbf{S}_{g(n)/2}$. So while we find bounds for the number of graphs that are connected and cellularly embeddable in $\mathbf{S}_{g(n)/2}$, we can not make a general statement on the size of the whole graph class from these as of yet. Unfortunately, the proof of the non-orientable case as in Section 3.3.2.2 does not seem to work in the unlabeled case.

3.6.2.1 Orientable Case

The number of unicellular rooted maps in the orientable surface of Euler genus $g(n)$ on n vertices is bounded above by $2^{4n+3g(n)} \cdot g(n)^{g(n)}$ from equation (3.28). In the unlabeled case, this is an upper bound on the number of possible connected 1-face simple unlabeled graph embeddings in the orientable surface of Euler genus $g(n)$ on n vertices. Once we

have found a connected unicellular embedded graph W, we add edges in the interior of the face bounded by the facial walk of the unique face of W so that no two edges in the interior cross. Recall that the embedding of W is cellular, so the unique face is homeomorphic to a disc. There are $2g(n)+2n-2$ vertices in the boundary walk around the unique face including repetitions, since every edge of W appears exactly twice in the boundary walk. From [32] we know that for any k there are at most $(2+3\sqrt{2})^k$ dissections of an k-gon, where the vertices are labelled $v_1, \ldots, v_k$ clockwise around the cycle on k vertices for k sufficiently large. For a given W, allowing double edges and loops, there are hence at most $(2+3\sqrt{2})^{2g(n)+2n-2}$ possible ways of filling the face, where $k = 2g(n)+2n-2$. This can be seen by taking the rooted edge of W as the starting edge v_1v_2 of the boundary cycle and noting that the labels $v_3, v_4, \ldots$ of the remaining vertices in the boundary cycle are determined by the embedding of W. Thus in total there are, for some constant $c_3 > 0$ at most

$$(2+3\sqrt{2})^{2g(n)+2n} \cdot 2^{4n+3g(n)} \cdot g(n)^{g(n)} \ll c_3^{n+g(n)} \cdot g(n)^{g(n)} \tag{3.105}$$

connected unlabelled graphs. Hence the number of graphs that are connected and have a cellular embedding in $\mathbf{S}_{g(n)/2}$ is at most $c_3^{n+g(n)} \cdot g(n)^{g(n)}$. We can obtain a bound on the number of graphs embeddable in any orientable surface or orientable Euler genus at most $g(n)$ by summing over all graphs cellularly embeddable in some surface of orientable Euler genus at most $h(n)$, where $h(n) \leqslant g(n)$. We hence have at most $g(n) \cdot c_3^{n+g(n)} \cdot g(n)^{g(n)}$ connected graphs embeddable in some surface of orientable Euler genus at most $g(n)$. To move from connected graphs with an embedding in an orientable surface of Euler genus at most $g(n)$ to all graphs with an embedding in an orientable surface of Euler genus at most $g(n)$, we use the following result:

Lemma 113. *[40] Let $\mathscr{A}$ be a bridge-addable class of graphs, and let $\mathscr{C}$ be the class of connected graphs in $\mathscr{A}$. For each positive integer n we have*

$$\left|\widetilde{\mathscr{A}_n}\right| \leqslant n^2 \cdot \left|\widetilde{\mathscr{C}_n}\right| \ .$$

Combining this result with the bound we found above, we see that there exists a constant $c_{15} > 0$ such that

$$\left|\widetilde{\mathscr{G}}_n^{g(n)}\right| \leqslant c_{15}^{n+g(n)} \cdot g(n)^{g(n)} \ . \tag{3.106}$$

It also follows directly that, as in the labelled case, all classes of unlabeled graphs of Euler genus any given $g(n) = O\left(\frac{n}{\log(n)}\right)$ are small.

3.6.2.2 Non-orientable Case

Unfortunately we do not have an exact counting formula for the number of unicellular maps on a non-orientable surface and the upper bound we have found in Section 3.3.2.2 is not easily amendable to work for unlabeled graphs. We have thus far not found good upper bounds on the number of unlabeled unicellular graphs on non-orientable surfaces of growing genus.

3.7 Further Open Problems and Considerations

3.7.1 Further Consideration - The Class of Graphs of Hereditarily Bounded Genus

Another idea for future consideration is the class of hereditary graphs, which simplifies some proofs of properties as can be shown below. So far this has only been considered on the side, but it might be interesting to investigate how this class relates to the non-hereditary classes and might aid in finding some further properties.

Definition. *We define the* ***orientable hereditary class of graphs*** $\mathscr{K}_n^{g(n)}$ *of Euler genus at most* $g(n)$ *to be the class of all graphs* G *on* n *vertices and embeddable in an orientable surface of Euler genus at most* $g(n)$ *and such that for any set of* m *vertices* $V \subseteq V(G)$*, the induced graph* $G[V]$ *is embeddable in an orientable surface of Euler genus at most* $g(m)$*.*

It turns out that some common properties of a graph chosen uniformly at random from the hereditary class are more easily deduced than in the general case. Note furthermore that when $g(n)$ satisfies $g(a+b) \geqslant g(a)+g(b)$ for all $a, b \in \mathbb{N}$, the class satisfies all addability conditions. We consider mostly the case when $g(n) \geqslant n$ in the following.

3.7.1.1 Number of Graphs of Hereditarily Bounded Genus

For some $k \in \mathbb{N}$, we wish to construct many graphs of hereditary Euler genus at most kn, that is graphs in the class $\mathscr{K}_n^{2kn}$. From $k+1$ graphs $G_1,\ldots,G_{k+1}$ in $\mathscr{P}_n$ we wish to construct many graphs in $\mathscr{K}_n^{2kn}$ with little double counting. Let the vertices in G_i be labeled from $(i-1)n+1$ to in. For each $j \in [k+1]$, we now identify the vertices $in+j$ for $i=0$ to $i=k$ with the resulting vertex labeled j to obtain the graph G'. The resulting graph G' is then in $\mathscr{K}_n^{2kn}$. This can be seen by noting that the disjoint union of all the G_i is still a planar graph. So if we add an edge between $in+j$ and $(i+1)n+j$ for each

$i \in [k-1]$ and each $j \in [k+1]$, call these edges "special" edges, we add in total at most kn edges. The resulting graph is hence embeddable in a surface of Euler genus at most $2kn$. By contracting all the "special" edges, we do not increase the genus of the graph and the resulting graph is G'. Further, deleting any set of vertices from the graph G' is the same as deleting this set of vertices from each of the planar graphs. The resulting graph has hence, by the same reasoning as before, genus $2k(n-m)$, where m is the number of vertices in the set of vertices we delete. The graph G' is hence of hereditary genus $2kn$ as required. Hence the Euler genus of G' is at most $2kn$. Furthermore, each ordered list of graphs $G_1, \ldots, G_{k+1}$ gives exactly one construction G'.

How often is each graph G' constructed? There are at most $3n$ edges in each of the graphs G_i, that is at most $3(k+1)n$ edges in total. Each edge in the new graph G' runs between u and v for some $u, v \in \{1, \ldots, n\}$. For each edge in the new graph, there must have been at least one edge between $in+u$ and $in+v$ for some $i \in \{0, \ldots, k\}$, but there can have been several such edges for different i.

Choose a construction uniformly at random from all constructions. For any possible edge between u and v in the resulting graph, the probability that the edge between $(i-1)+u$ and $(i-1)+v$ is present in the planar graph P_i is at most $3n/\binom{n}{2} \sim 6/n$. This is the case since planarity is retained under rotation of the vertices and as a result every possible edge on n vertices has equal probability of being present in a graph chosen uniformly at random from all planar graphs. Thus, the probability that the edges between $(i-1)+v$ and $(i-1)+u$ and between $(j-1)+v$ and $(j-1)+u$ are present simultaneously in P_i and P_j is at most

$$(6/n)^2 \ .$$

We call an edge uv in a construction a "double" edge if there exists at least two graphs P_i and P_j such that the edges $[(i-1)+u][(i-1)+v]$ and $[(j-1)+u][(j-1)+v]$ are both present in the graphs P_i and P_j respectively. Thus, letting X be the number of "double edges" in a construction chosen uniformly at random from all constructions, we have that the expected number of "double edges" is

$$\begin{aligned}
\mathbb{E}[X] &\leqslant \binom{n}{2} \cdot \binom{k+1}{2} \left(\frac{6}{n}\right)^2 \\
&\sim \frac{n^2}{2} \cdot \frac{k^2+k}{2} \cdot \frac{6^2}{n^2} \\
&\to 9(k^2+k) \\
&< 10k^2 \ , \qquad (3.107)
\end{aligned}$$

as $n \to \infty$. Thus at least half of all the graphs constructed have at most $20k^2$ double edges and no triple or higher edges. For each non-double edge e in a construction, of

the corresponding edges in the planar graphs exactly one must be present in some P_i. There are therefore $(k+1)$ possibilities for a non-double edge to be constructed. For each double-edge, at least two of the corresponding edges in the planar graphs must be present, implying that there are at most $2^{k+1}-(k+2)$ possible ways of constructing a double edge. In total there are at most $3(k+1)n$ edges present in the constructed graph, meaning that half of the graphs G' are constructed at most

$$(k+1)^{(3n-18)(k+1)} \cdot 2^{18(k+1)^2}$$

times. We thus see, since by [34] $|\mathscr{P}_n| \sim g \cdot n^{-\frac{7}{2}} \cdot \gamma_l^n \cdot n!$, that:

$$\begin{aligned}
\left|\mathscr{K}_n^{2kn}\right| &\geqslant \frac{1}{2}\left(|\mathscr{P}_n|\right)^{(k+1)} \cdot \frac{1}{(k+1)^{(3n-18)(k+1)} \cdot 2^{18(k+1)^2}} \\
&\geqslant \frac{1}{2}\left(g \cdot n^{-\frac{7}{2}} \cdot \gamma_l^n \cdot n!\right)^{(k+1)} \cdot (k+1)^{-(3n-18)(k+1)} \cdot 2^{-18(k+1)^2} \\
&= \frac{1}{2} \cdot g^{(k+1)} \cdot n^{-\frac{7(k+1)}{2}} \cdot \gamma_l^{(k+1)n} \cdot (n!)^{(k+1)} \cdot (k+1)^{-(3n-18)(k+1)} \cdot 2^{-18(k+1)^2} \\
&\geqslant \frac{1}{2} \cdot n! \cdot \frac{(k+1)^{18(k+1)} \cdot g^{(k+1)} \cdot n^{-\frac{7(k+1)}{2}}}{2^{18(k+1)^2}} \cdot \left(\frac{n}{e}\right)^{kn} \cdot \gamma_l^{(k+1)n} \cdot 2^{3(k+1)n} \\
&\geqslant \frac{1}{2} \cdot n! \cdot \frac{(k+1)^{18(k+1)} \cdot g^{(k+1)} \cdot n^{-\frac{7(k+1)}{2}}}{2^{18(k+1)^2}} \cdot \left(\frac{2^{3(k+1)} \cdot \gamma_l^{(k+1)} \cdot n^k}{e^k}\right)^n . \qquad (3.108)
\end{aligned}$$

So

$$\left|\mathscr{K}_n^{2kn}\right| \geqslant n! \cdot \frac{(k+1)^{18(k+1)} \cdot g^{(k+1)} \cdot n^{-\frac{7(k+1)}{2}}}{2 \cdot 2^{18(k+1)^2}} \cdot \left(\frac{2^{3(k+1)} \cdot \gamma_l^{(k+1)} \cdot n^k}{e^k}\right)^n ,$$

or alternatively,

$$\begin{aligned}
\left(\frac{\left|\mathscr{K}_n^{2kn}\right|}{n!}\right)^{\frac{1}{n}} &\geqslant \left(\frac{(k+1)^{18(k+1)} \cdot g^{(k+1)} \cdot n^{-\frac{7(k+1)}{2}}}{2 \cdot 2^{18(k+1)^2}} \cdot \left(\frac{2^{3(k+1)} \cdot \gamma_l^{(k+1)} \cdot n^k}{e^k}\right)^n\right)^{\frac{1}{n}} \\
&= \left(\frac{(k+1)^{18(k+1)} \cdot g^{(k+1)} \cdot n^{-\frac{7(k+1)}{2}}}{2 \cdot 2^{18(k+1)^2}}\right)^{\frac{1}{n}} \cdot \frac{2^{3(k+1)} \cdot \gamma_l^{(k+1)} \cdot n^k}{e^k} \\
&= (1+o(1)) \cdot \frac{2^{3(k+1)} \cdot \gamma_l^{(k+1)}}{e^k} \cdot n^k . \qquad (3.109)
\end{aligned}$$

Hence, all classes of graphs $\mathscr{K}_n^{g(n)}$ for some $g(n) \geqslant 2n$ are over-large. Furthermore, we get a bound on the size of $\mathscr{K}_n^{g(n)}$ by letting $k = \left\lfloor \frac{g(n)}{2n} \right\rfloor$

$$\left|\mathscr{K}_n^{g(n)}\right| \geqslant (1+o(1))^n \cdot \frac{2^{3\left(\left\lfloor \frac{g(n)}{2n} \right\rfloor+1\right)\cdot n} \cdot \gamma_l^{\left(\left\lfloor \frac{g(n)}{2n} \right\rfloor+1\right)\cdot n}}{c^{\left\lfloor \frac{g(n)}{2n} \right\rfloor}} \cdot n^{\left\lfloor \frac{g(n)}{2n} \right\rfloor \cdot n} \cdot n! . \qquad (3.110)$$

3.7.1.2 Pendant Vertices

We now want to show that for any genus function satisfying $g(a+b) \geqslant g(a)+g(b)$ for all $a, b \in \mathbb{N}$, a graph chosen uniformly at random from $\mathscr{K}_n^{g(n)}$ has with high probability few pendant vertices. Let $g(n) \geqslant 2n$ and $\alpha(n)$ be a function of n that tends to infinity as $n \to \infty$ and $\alpha(n) \leqslant n/2$. We define

$$\begin{aligned}\mathscr{A}_n &= \mathscr{K}_n^{g(n)} \\ \mathscr{B}_n &= \{G \in \mathscr{H}_n^{g(n)} \mid G \text{ has } \geqslant \alpha(n) \text{ pendant vertices}\} .\end{aligned}$$

Let $G \in \mathscr{B}_n$. We wish to construct many graphs in $\mathscr{A}_n$ from G with little double counting. By $v_1, v_2, \ldots, v_k$ denote the $k \geqslant \alpha(n)$ pendant vertices of G. The graph $G \setminus \{v_1, \ldots, v_{\alpha(n)}\}$ then has genus at most $g(n-2\alpha(n)) \leqslant g(n)-2\alpha(n)$ by the hereditary property of G and superadditivity. This implies that the original graph G also has genus at most $g(n-\alpha(n)) \leqslant g(n)-2\alpha(n)$ since adding a pendant vertex does not increase the genus. We can hence increase the genus by at least $2\alpha(n)$ by adding edges to G, such that the graph constructed G' still has genus at most $g(n)$. For each $v_1, \ldots, v_{\alpha(n)}$, add an extra edge connecting it to any of the other $n-\alpha(n)$ vertices. There are $(n-\alpha(n))^{\alpha(n)}$ ways of doing so. This will increase the Euler genus by at most two for each vertex v_i, thus by at most $2\alpha(n)$ in total. Note also that by adding these edges, the hereditary genus property of the graph is not destroyed, meaning that the new graph G' still has hereditary genus $g(n)$. So $G \in \mathscr{A}_n$.

How often is each G' constructed? Guess the $\alpha(n)$ vertices that were originally pendant vertices, there are at most $\binom{n}{\alpha(n)}$ choices. During the construction, we have added exactly one edge to each of these vertices. There are at most $\binom{2\alpha(n)}{\alpha(n)} \leqslant 2^{\alpha(n)}$ ways of deleting exactly one edge from each vertex. Thus, each G' is constructed at most

$$\binom{n}{\alpha(n)} \cdot 2^{\alpha(n)}$$

times. This gives that, since for n large enough $1-\frac{\alpha(n)}{n} \geqslant \frac{1}{2e}$,

$$\begin{aligned}|\mathscr{A}_n| &\geqslant |\mathscr{B}_n| \cdot \frac{\left(\left(1-\frac{\alpha(n)}{n}\right)n\right)^{\alpha(n)}}{\binom{n}{\alpha(n)} \cdot 2^{\alpha(n)}} \\ &\geqslant |\mathscr{B}_n| \cdot \frac{\left(1-\frac{\alpha(n)}{n}\right)^{\alpha(n)} \cdot n^{\alpha(n)}}{\left(\frac{n \cdot e}{\alpha(n)}\right)^{\alpha(n)} \cdot 2^{\alpha(n)}} \\ &\geqslant |\mathscr{B}_n| \cdot \left(\frac{1}{2e}\right)^{2\alpha(n)} \cdot \alpha(n)^{\alpha(n)} \\ &\geqslant |\mathscr{B}_n| \cdot c^{\alpha n} \end{aligned} \tag{3.111}$$

for some constant $c > 0$ and n large enough since for all n large enough there exists a constant c such that

$$\frac{\alpha(n)}{2^2 \cdot e^2} > c$$

as $\alpha(n) \to \infty$ as $n \to \infty$.

Therefore

$$\mathbb{P}(G \in \mathscr{B}_n) \leqslant c^{-\alpha(n)} \to 0 \tag{3.112}$$

as $n \to \infty$.

Thus for all $\epsilon > 0$, there exists a constant $k \in \mathbb{N}$ such that with probability at least $1 - \epsilon$ a graph G chosen uniformly at random from $\mathscr{H}_n^{g(n)}$ has at most k pendant vertices.

3.7.2 Further Open Problems

Before we started our work on random graphs on non-constant surfaces there were no known results about the sizes of these graph classes or their properties. We have hence managed to take a big step towards understanding these classes better. However, answering some questions brought up many more questions whose answers we do not know yet. In this section, we attempt to give a brief overview of some of the open problems we have come across.

We have thus far managed to find some upper and lower bounds on the sizes of $\left|\mathscr{G}_n^{g(n)}\right|$ and $\left|\mathscr{H}_n^{g(n)}\right|$ for all suitable functions $g(n)$. Are any of these bounds tight? We expect there to be the planar graph growth constant for sufficiently smooth $g(n)$ up to genus $g(n) = o(n/\log(n))$ and another growth constant for any $g(n) = \Theta(n/\log(n))$. How does the growth constant, if it exists, depend on α for $g(n) = \alpha n/\log(n)$? Is it possible to actually show this existence for genus $g(n) = \omega(n/\log^3(n))$? Furthermore, for large genus, that is genus $g(n) = \omega(n)$, what is the leading term in the growth rate? Currently the lower bounds we found are of order $\Theta\left((n^2/g(n))^{g(n)}\right)$, while the upper bounds are of order $g(n)^{g(n)}$. We expect the reality to lie closer to the upper than the lower bound and in fact conjecture that the actual rate of growth is of order $g(n)^{g(n)}$ as long as $g(n)$ is not too large.

Conjecture 114. *There is a growth constant equal to the planar graph growth constant for all $\mathscr{G}_n^{g(n)}$ and $\mathscr{H}_n^{g(n)}$ as long as $g(n) = o\left(\frac{n}{\log(n)}\right)$.*

Conjecture 115. *There is a growth constant $\gamma = \gamma(c) > \gamma_{\mathscr{P}}$ for all $\mathscr{G}_n^{g(n)}$ and $\mathscr{H}_n^{g(n)}$ for sufficiently smooth $g(n)$ and where $g(n) = \Theta\left(\frac{cn}{\log(n)}\right)$.*

While we have managed to make some statements about the growth of the graph classes $\mathscr{G}_n^{g(n)}$ and $\mathscr{H}_n^{g(n)}$, we still do not know much about the increase in the size of these graph classes when moving from n to $n+1$ vertices. It is obvious that one has to distinguish between the case when the Euler genus does not increase and when the Euler genus increases by at least one. We call the latter a jump step. In Lemma 76 we have managed to find some lower bounds on the increase from $\left|\mathscr{G}_n^{g(n)}\right|$ to $\left|\mathscr{G}_{n+1}^{g(n+1)}\right|$. However, these bounds are by no means tight. We can improve on the bounds found in Lemma 76 with the $(n+1)$ in Lemma 76 replaced by $2n$ as seen in the following Lemma, however these bounds still do not appear to be tight. What is the correct increase in size in a jump step?

Lemma 116. *For any genus $h \geqslant 0$ and $n \in \mathbb{N}$ we have that*

$$\left|\mathscr{G}_{n+1}^h\right| \geqslant 2n \cdot \left|\mathscr{G}_n^h\right| \ . \tag{3.113}$$

Proof. Note that in Lemma 76, we have already proven a bound of similar kind. We prove this new bound by another double counting argument. Taking a graph G on n vertices in $\mathscr{G}_n^h$, we wish to attach vertex $n+1$ to the graph without increasing the genus. We can do so by attaching vertex $n+1$ as either a pendant vertex adjacent to any of the n already existent vertices, or as an isolated vertex. Furthermore, we can attach vertex $n+1$ to both endvertices of any of the already existent edges without increasing the genus. If the graph G is connected, there are at least $n-1$ choices for doing so. The situation is slightly different when the graph consists of k components, we however show that even in this case there are still $2n$ choices. Suppose the graph is consisting of k parts, then we can attach vertex $n+1$ to the endvertices of any edge or incident to exactly one vertex in said part or not incident to any vertices in this part independently of all other parts. Thus, if there are k parts with n_i vertices in the i^{th} part, we have $(n_i - 1) + n_i + 1 = 2n_i$ choices of how to attach (or not attach) vertex $n+1$ in the i^{th} part, independently of all other parts. In total we hence have at least

$$\prod_{i=0}^{k} (2n_i) \geqslant 2n$$

choices. Since each graph on $n+1$ vertices and with the same genus is constructed exactly once (we uniquely return to G by deleting vertex $n+1$), we can deduce the required result. ■

We conjecture that the correct increase in graph class size when adding a vertex but not increasing the genus is roughly an extra factor of $n \cdot \gamma_{\mathscr{G}}$.

Conjecture 117. *For any $\epsilon > 0$ there is an n_0 such that for each $n \geqslant n_0$ and each $h \geqslant 0$*

$$\left|\mathscr{G}_{n+1}^{h}\right| \geqslant (1-\epsilon) \cdot n \cdot \gamma_{\mathscr{P}} \cdot \left|\mathscr{G}_{n}^{h}\right| \ .$$

We can also improve the bound from Lemma 116 for very large genus.

Lemma 118. *Note that for any $g(n) = \omega(n)$ such that $g(n) = o(n^2)$, $n \in \mathbb{N}$ and $p \in (0,1)$ we have that*

$$\left|\mathscr{G}_{n+1}^{g(n)}\right| \geqslant p \cdot g(n) \cdot \left|\mathscr{G}_{n}^{g(n)}\right| .$$

Proof. We now wish to strengthen the above bound for large genus. We prove this bound by a double counting argument. Note that when adding a vertex with label $n+1$ to an n-vertex graph embedded to a surface we can join the new vertex to the endvertices of any edge without increasing the genus of the embedding. Alternatively, we can attach the new vertex to any existing vertex as a pendant vertex or to no vertex at all without increasing the genus. By Theorem 104 any graph chosen uniformly at random from $\mathscr{G}_n^{g(n)}$ has, as long as $g(n) = o(n^2)$, with probability at least $1-\epsilon$ for n sufficiently large at least $g(n)+n$ edges. Choosing $\epsilon = 1-p$ we see that there are for at least p graphs in $\mathscr{G}_n^{g(n)}$ at least $g(n)+n \geqslant g(n)$ choices for where to attach vertex $n+1$ adjacent to both endvertices of an edge. Further, if we attach $n+1$ to a vertex as a pendant vertex, there are n choices for this and one choice for attaching it to no vertex at all. Thus in total there are, for a fraction of at least p graphs in $\mathscr{G}_n^{g(n)}$, at least $g(n)+n+1 = n+g(n)+1$ constructions G'. For the remaining fraction of $(1-p)$ graphs in $\mathscr{G}_n^{g(n)}$, there are at least $n+1$ possible constructions G', by attaching $n+1$ as a pendant vertex, or not attaching it at all. Note that by deleting vertex $n+1$ from G' we uniquely get back to the graph G. Thus

$$\left|\mathscr{G}_{n+1}^{g(n)}\right| \geqslant (n+g(n)+1)p\left|\mathscr{G}_{n}^{g(n)}\right| + (1-p)n\left|\mathscr{G}_{n}^{g(n)}\right| \ ,$$

from which the required bound follows. ■

Can we find a corresponding upper bound? And what is the actual rate of increase? Is the growth of the graph classes smooth?

In the final main part of the thesis, we have found some common properties of a graph chosen uniformly at random from the graph classes $\mathscr{G}_n^{g(n)}$ and $\mathscr{H}_n^{g(n)}$. Much however remains open about the properties. In terms of pendant appearances of vertices or other graphs, we know that there exists with high probability linearly many such pendant copies with high probability as long as a growth constant exists. Furthermore, as long as the graph class is small, the radius of convergence of the subclass of graphs having fewer than linearly many pendant appearances is strictly larger than the radius of convergence of the

entire class. If the class satisfies a certain smoothness criterion this would imply that there are with high probability linearly many pendant appearances. Thus far we have however not managed to prove this result. For this, as well as all other properties we have looked at, we have been unable to make any statements for the range $\omega(n/\log(n)) = g(n) = o(n)$.

In terms of connectivity, we know that the probability of connectedness is bounded away from zero for all bridge-addable classes. Further, as long as growth constant exists, the probability of connectedness is also in the limit bounded away from one. Furthermore, as long as $g(n) = \omega(n)$, the probability of connectedness is with high probability equal to one. When does the transition between being bounded away from one and being with high probability equal to one occur? What happens in the range $\omega(n/\log(n)) = g(n) = o(n)$?

Finally, looking at the maximum degree, we know that as long as a growth constant exists, the maximum degree is with high probability of order $\Theta(\log(n))$. Once the genus function satisfies $g(n) = \omega(n\log(n))$, the maximum degree is at least $\Theta(g(n)/n)$ with high probability, since this is with high probability the average degree of a graph chosen uniformly at random from the class. We have however been unable to make any statements about upper bounds on the maximum degree in this range or any statement about the maximum degree in the range $\omega(n/\log(n)) = g(n) = o(n\log(n))$. We conjecture that the maximum degree is $\Theta(\log(n))$ for genus $g(n) = o(n)$ and of order $\Theta\left(\frac{g(n)}{n}\log(n)\right)$, as in the Erdős–Rényi random graph model for large average degree, for $g(n) = \omega(n)$.

Bibliography

[1] Andrew D. Barbour, Michal Karoński, and Andrzej Ruciński, A central limit theorem for decomposable random variables with applications to random graphs, *Journal of Combinatorial Theory Series B* 47.2 (1989), 125–145.

[2] Ross Berkowitz, A local limit theorem for cliques in $G(n, p)$, *http://arxiv.org/abs/1811.03527* (2018).

[3] Béla Bollobás, Threshold functions for small subgraphs, *Mathematical Proceedings of the Cambridge Philosophical Society* 90.2, Cambridge University Press (1981), 197–206.

[4] Béla Bollobás, *Random Graphs*, First Edition, Academic Press, London 1985, Second Edition, Cambridge University Press, 2001.

[5] Rick Durrett, *Probability: Theory and Examples*, Cambridge University Press (2010).

[6] Paul Erdős, Graph Theory and Probability, *Canadian Journal of Mathematics* 11 (1959), 34–38

[7] Paul Erdős and Alfréd Rényi, On random graphs I, *Publicationes Mathematicae* 6 (1959) 290–297.

[8] Paul Erdős and Alfréd Rényi, On the evolution of random graphs, *Publications of the Mathematical Institute of the Hungarian Academy of Sciences, Series A* 5.1 (1960), 17–61.

[9] Pál Erdős and Alfréd Rényi, On the strength of connectedness of a random graph, *Acta Mathematica Academiae Scientiarum Hungaricae* 12 (1961), 261–267.

[10] Pál Erdős and Alfréd Rényi, On random matrices, *Publications of the Mathematical Institute of the Hungarian Academy of Sciences, Series A* 8 (1964), 455–461.

[11] Pál Erdős and Alfréd Rényi, On the existence of a factor of degree one of a connected random graph, *Acta Mathematica Academiae Scientiarum Hungaricae* 17 (1966), 359–368.

[12] William Feller, *An Introduction to Probability Theory and its Applications*, Vol. II. Second edition, John Wiley & Sons, Inc., New York-London-Sydney (1971).

[13] Edgar N. Gilbert, Random graphs, *Annals of Mathematical Statistics* 30 (1959), 1141–1144.

[14] Justin Gilmer and Swastik Kopparty, A local central limit theorem for triangles in a random graph, *Random Structures and Algorithms* 48.4 (2016), 732–750.

[15] Wassily Hoeffding, Probability inequalities for sums of bounded random variables, *Journal of the American Statistical Association* 58.301 (1963), 13–30.

[16] Svante Janson, Tomasz Łuczak and Andrzej Ruciński, An exponential bound for the probability of nonexistence of a specified subgraph in a random graph, *Proceedings of Random Graphs' 87*, Wiley, Chichester (1990), 73–87.

[17] Michał Karoński and Andrezj Ruciński, On the number of strictly balanced subgraphs of a random graph, *Graph theory Lagów 1981, Lecture Notes in Mathematics* 1018, Springer, Berlin, Heidelberg (1983), 79–83.

[18] Krzysztof Nowicki, Asymptotic normality of graph statistics, *Journal of Statistical Planning and Inference* 21.2 (1989), 209–222.

[19] Michael Molloy and Bruce Reed, Graph colouring and the probabilistic method, *Algorithms and Combinatorics* 23, Springer, Berlin, Heidelberg, (2013).

[20] Andrzej Ruciński, When are small subgraphs of a random graph normally distributed? *Probability Theory and Related Fields* 78.1 (1988), 1–10.

[21] Jason P. Bell, Edward A. Bender, Peter J. Cameron and L. Bruce Richmond, Asymptotics for the probability of connectedness and the distribution of number of components, *The Electronic Journal of Combinatorics* 7.1 (2000), R33,22.

[22] Edward A. Bender, Zhicheng Gao and Nicholas C. Wormald, The number of labeled 2-connected planar graphs, *Electronic Journal of Combinatorics* 9.1 (2002), R43, 13.

[23] Olivier Bernardi and Guillaume Chapuy, Counting unicellular maps on non-orientable surfaces, *Advances in Applied Mathematics* 47.2 (2011), 259–275.

[24] Manuel Bodirsky, Omer Giménez, Mihyun Kang and Marc Noy, On the number of series parallel and outerplanar graphs, *Proceedings of the European Conference on Combinatorics, Graph Theory, and Applications (EuroComb 2005)*, DMTCS Proceedings Series Volume AE (2005), 383–388.

[25] Nicolas Bonichon, Cyril Gavoille, Nicolas Hanusse, Dominique Poulalhon and Gilles Schaeffer, Planar graphs, via well-orderly maps and trees, *Graphs and Combinatorics* 22.2 (2006), 185–202.

[26] Guillaume Chapuy Éric Fusy, Omer Giménez, Bojan Mohar and Marc Noy, Asymptotic enumeration and limit laws for graphs of fixed genus, *Journal of Combinatorial Theory, Series A* 118.3 (2011), 748–777.

[27] Éric Colin De Verdière, Alfredo Hubard and Arnaud de Mesmay, Discrete systolic inequalities and decompositions of triangulated surfaces, *Discrete and Computational Geometry* 53.3 (2015), 587–620.

[28] Alain Denise, Marcio Vasconcellos and Dominic JA Welsh, The random planar graph, *Congressus Numerantium* 113 (1996), 61–79.

[29] Krzystof Diks, Hristo N. Djidjev, Ondrej Sykora and Imrich Vrto, *Edge separators for planar graphs and their applications*, *Mathematical Foundations of Computer Science 1988, Proceedings of* 13^{th} *Symposium, Lecture Notes Computer Science* 324 (1988), 280–290.

[30] Hristo N. Djidjev and Shankar M. Venkatesan, Planarization of graphs embedded on surfaces, *Proceedings of the 21st International Workshop on Graph-Theoretic Concepts in Computer Science* (1995), 62–72.

[31] J. R. Fiedler, J. P. Huneke, R. B. Richter and N. Robertson, Computing the orientable genus of projective graphs, *Journal of Graph Theory* 20.3 (1995), 297–308.

[32] Philippe Flajolet and Marc Noy, Analytic combinatorics of non-crossing configurations, *Discrete Mathematics* 204.1-3 (1999), 203–229.

[33] John R. Gilbert, Joan P. Hutchinson and Robert Endre Tarjan, A separator theorem for graphs of bounded genus, *Journal of Algorithms* 5.3 (1984), 391-407.

[34] Omer Giménez and Marc Noy. Asymptotic enumeration and limit laws of planar graphs, *Journal of the American Mathematical Society* 22.2 (2009), 309–329.

[35] Omer Giménez and Marc Noy, The number of planar graphs and properties of random planar graphs, *International Conference on Analysis of Algorithms, Papers from the conference*, Barcelona, Spain (2005) 147–156.

[36] Frank Harary, Paul C. Kainen, Allen J. Schwenk and Arthur T. White, A maximal toroidal graph which is not a triangulation, *Mathematica Scandinavica* 33 (1973), 108–112.

[37] Evgeniy Krasko, Igor Labutin and Alexander Omelchenko, Enumeration of 3-regular one-face maps on orientable or non-orientable surface up to all symmetries, *http://arXiv.org/abs/1901.06591* (2019).

[38] Richard J. Lipton and Robert Endre Tarjan, A separator theorem for planar graphs, *SIAM Journal on Applied Mathematics* 36 (1979), 177–189.

[39] Colin McDiarmid, Connectivity for random graphs from a weighted bridge-addable class, *The Electronic Journal of Combinatorics* 19.4 (2012) 53–72.

[40] Colin McDiarmid. Unpublished Manuscript.

[41] Colin McDiarmid, Random graphs from a minor-closed class, *Combinatorics, Probability and Computing* 18.4 (2009), 583–599.

[42] Colin McDiarmid, Random graphs on surfaces, *Journal of Combinatorial Theory, Series B* 98.4 (2008), 778–797.

[43] Colin McDiarmid and Bruce Reed, On the maximum degree of a random planar graph, *Combinatorics, Probability and Computing* 17.4 (2008), 591–601.

[44] Colin McDiarmid and Marc Noy. *Asymptotic enumeration, random planar graphs, and beyond* Unpublished manuscript

[45] Colin McDiarmid, Angelika Steger and Dominic JA Welsh, Random planar graphs, *Journal of Combinatorial Theory, Series B* 93.2 (2005), 187–205.

[46] Colin McDiarmid, Angelika Steger and Dominic JA Welsh, Random graphs from planar and other addable classes, *Topics in Discrete Mathematics*, Springer, Berlin, Heidelberg, (2006) 231–246.

[47] Colin McDiarmid and David R. Wood, Edge-Maximal Graphs on Surfaces, *Canadian Journal of Mathematics* 70.4 (2018) 925–942.

[48] Bojan Mohar and Carsten Thomassen, *Graphs on surfaces*. Baltimore, MD, Johns Hopkins University Press (2001).

[49] J. W. Moon, Counting Labelled Trees, *Canadian Mathematical Monographs 1, Canadian Mathematical Congress* (1970).

[50] Gerhard Ringel, Bestimmung der Maximalzahl der Nachbargebiete auf nicht–orientierbaren Flächen, *Mathematische Annalen* 127.1 (1954), 181–214.

[51] Gerhard Ringel and John WT Youngs, Solution of the Heawood map-coloring problem, *Proceedings of the National Academy of Sciences* 60.2 (1968), 438–445.

[52] N. Robertson, D. Sanders, P.D. Seymour and R. Thomas, The four-colour theorem, *Journal of Combinatorial Theory, Series B* 70.1 (1997), 2–44.

[53] Ondrej Sýkora and Imrich Vrto, Edge separators for graphs of bounded genus with applications, *Theoretical Computer Science* 112.2 (1993), 419–429.

[54] Carsten Thomassen, The Jordan-Schonflies Theorem and the Classification of Surfaces, *The American Mathematical Monthly* 99.2 (1992), 116–131.

[55] Timothy RS. Walsh and Alfred B. Lehman, Counting rooted maps by genus. I, *Journal of Combinatorial Theory, Series B* 13 (1972), 192–218.

[56] E.M. Wright, A relationship between two sequences. I, *Proceedings of the London Mathematical Society. Third Series* 17 (1967), 296–304.

[57] E. M. Wright, A relationship between two sequences. III, *Journal of the London Mathematical Society* 43 (1968), 720–724.

[58] Mihalis Yannakakis, Node- and edge-deletion NP-complete problems, *Proceedings of the Tenth Annual ACM Symposium on Theory of Computing* (1978), 253–264.